

;; INFORMATION FOR SEQ ID NO: 14:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 10  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: unknown  
US-08-986-106-14

Query Match 100.0%; Score 18; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
1111  
DB 2 NLAA 5

RESULT 9  
US-08-986-106-15  
; Sequence 15, Application US/08986106  
; Patent No. 6177410

GENERAL INFORMATION:

APPLICANT: HOLT, JEFFREY T.

APPLICANT: JENSEN, ROY A.

APPLICANT: KING, MARY-CLAIRE

APPLICANT: STEINER, MITCHELL S.

APPLICANT: ROBINSON-BENION, CHERYL L.

APPLICANT: THOMPSON, MARILYN E.

TITLE OF INVENTION: THERAPEUTIC METHODS FOR

TITLE OF INVENTION: PROSTATE CANCER

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARLES A. TAYLOR, JR.

STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER

STREET: BOULEVARD

CITY: DURHAM

STATE: NORTH CAROLINA

COUNTRY: USA

ZIP: 27707

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage

COMPUTER: IBM PC/XT/AT compatible

OPERATING SYSTEM: Windows 3.1

SOFTWARE: WORD PERFECT 6.1 and ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/986,106

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/603,753

FILING DATE: 20 FEB 1996

ATTORNEY/AGENT INFORMATION:

NAME: ARLES A. TAYLOR, JR.

REGISTRATION NUMBER: 39,395

REFERENCE/DOCKET NUMBER: 1242/3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919) 493-8000

TELEFAX: (919) 419-0383

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 10

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

US-08-986-106-15

DB 2 NLAA 5  
1111

RESULT 10  
US-08-986-106-16  
; Sequence 16, Application US/08986106  
; Patent No. 6177410

GENERAL INFORMATION:

APPLICANT: HOLT, JEFFREY T.

APPLICANT: JENSEN, ROY A.

APPLICANT: KING, MARY-CLAIRE

APPLICANT: STEINER, MITCHELL S.

APPLICANT: ROBINSON-BENION, CHERYL L.

APPLICANT: THOMPSON, MARILYN E.

TITLE OF INVENTION: THERAPEUTIC METHODS FOR

TITLE OF INVENTION: PROSTATE CANCER

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARLES A. TAYLOR, JR.

STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER

STREET: BOULEVARD

CITY: DURHAM

STATE: NORTH CAROLINA

COUNTRY: USA

ZIP: 27707

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage

COMPUTER: IBM PC/XT/AT compatible

OPERATING SYSTEM: Windows 3.1

SOFTWARE: WORD PERFECT 6.1 and ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/986,106

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/603,753

FILING DATE: 20 FEB 1996

ATTORNEY/AGENT INFORMATION:

NAME: ARLES A. TAYLOR, JR.

REGISTRATION NUMBER: 39,395

REFERENCE/DOCKET NUMBER: 1242/3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919) 493-8000

TELEFAX: (919) 419-0383

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 10

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

US-08-986-106-16

Query Match 100.0%; Score 18; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
1111  
DB 2 NLAA 5

RESULT 11

US-08-258-639A-5

; Sequence 5, Application US/08258639A

; Patent No. 5681733

GENERAL INFORMATION:

APPLICANT: Su, Hongsheng

APPLICANT: Blain, Francoise

APPLICANT: Bennett, Clark

APPLICANT: Gu, Kangfu

APPLICANT: Zimmermann, Joseph

OY 1 NLAA 4

VOLUME:  
PAGES:  
DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 18:  
US-09-099-753-18

Query Match 100.0%; Score 18; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4  
1111  
DB 2 NLAA 5

RESULT 7  
US-09-099-753-19  
Sequence 19, Application US/09099753  
Patent No. 6149903  
GENERAL INFORMATION:  
APPLICANT: HOLT, JEFFREY T.  
APPLICANT: JENSEN, ROY A.  
APPLICANT: PAGE, DAVID L.  
APPLICANT: KING, MARY-CLAIRE  
APPLICANT: SZABO, CSILLA I.  
APPLICANT: JETTTON, THOMAS L.  
APPLICANT: ROBINSON-BENION, CHERYL L.  
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2  
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARLES A. TAYLOR, JR.  
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER  
STREET: BOULEVARD  
CITY: DURHAM  
STATE: NORTH CAROLINA  
COUNTRY: USA  
ZIP: 27707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
COMPUTER: IBM PC/XT/AT compatible  
OPERATING SYSTEM: Windows 3.1  
SOFTWARE: WORD PERFECT 6.1 and ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/099,753  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/603,753  
FILING DATE: 20 FEB 1996  
APPLICATION NUMBER: U.S. 08/373,799  
FILING DATE: 17 JAN 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: ARLES A. TAYLOR, JR.  
REGISTRATION NUMBER: 39,395  
REFERENCE/DOCKET NUMBER: 1242/2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 493-8000  
TELEFAX: (919) 419-0383  
TELEX:  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: unknown  
HYPOTHETICAL: no  
ANTI-SENSE: no  
ORIGINAL SOURCE:

ORGANISM: Mouse  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
TISSUE TYPE:  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
FEATURE:  
NAME/KEY: Chromogranin B Granin Sequence  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
PAGES:  
DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 19:  
US-09-099-753-19  
Query Match 100.0%; Score 18; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 NLAA 4  
1111  
DB 2 NLAA 5

RESULT 8  
US-08-986-106-14  
Sequence 14, Application US/08986106  
Patent No. 6177410  
GENERAL INFORMATION:  
APPLICANT: HOLT, JEFFREY T.  
APPLICANT: JENSEN, ROY A.  
APPLICANT: KING, MARY-CLAIRE  
APPLICANT: STEINER, MITCHELL S.  
APPLICANT: ROBINSON-BENION, CHERYL L.  
TITLE OF INVENTION: THERAPEUTIC METHODS FOR  
TITLE OF INVENTION: PROSTATE CANCER  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARLES A. TAYLOR, JR.  
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER  
STREET: BOULEVARD  
CITY: DURHAM  
STATE: NORTH CAROLINA  
COUNTRY: USA  
ZIP: 27707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage  
COMPUTER: IBM PC/XT/AT compatible  
OPERATING SYSTEM: Windows 3.1  
SOFTWARE: WORD PERFECT 6.1 and ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/986,106  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/603,753  
FILING DATE: 20 FEB 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: ARLES A. TAYLOR, JR.  
REGISTRATION NUMBER: 39,395  
REFERENCE/DOCKET NUMBER: 1242/3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 493-8000  
TELEFAX: (919) 419-0383

NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARLES A. TAYLOR, JR.  
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER  
CITY: DURHAM  
STATE: NORTH CAROLINA  
COUNTRY: USA  
ZIP: 27707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
COMPUTER: IBM PC/XT/AT compatible  
OPERATING SYSTEM: Windows 3.1  
SOFTWARE: WORD PERFECT 6.1 and ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/099,753  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/603,753  
FILING DATE: 20 FEB 1996  
APPLICATION NUMBER: U.S. 08/373,799  
FILING DATE: 17 JAN 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: ARLES A. TAYLOR, JR.  
REGISTRATION NUMBER: 39,395  
REFERENCE/DOCKET NUMBER: 1242/2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 493-8000  
TELEFAX: (919) 419-0383  
TELEX:  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: no  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: Human  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
TISSUE TYPE:  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
FEATURE:  
NAME/KEY: Chromogranin B Granin Sequence  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
PAGES:  
DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 17:  
US-09-099-753-17

Query Match 100.0%; Score 18; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 NLAA 4  
DB 2 NLAA 5

RESULT 6  
US-09-099-753-18  
Sequence 18, Application US/09099753  
Patent No. 6149903  
GENERAL INFORMATION:  
APPLICANT: HOLT, JEFFREY T.  
APPLICANT: JENSEN, ROY A.  
APPLICANT: PAGE, DAVID L.  
APPLICANT: KING, MARY-CLAIRE  
APPLICANT: SZABO, CSILLA I.  
APPLICANT: JETTON, THOMAS L.  
APPLICANT: ROBINSON-BENION, CHERYL L.  
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2  
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON  
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARLES A. TAYLOR, JR.  
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER  
CITY: DURHAM  
STATE: NORTH CAROLINA  
COUNTRY: USA  
ZIP: 27707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
COMPUTER: IBM PC/XT/AT compatible  
OPERATING SYSTEM: Windows 3.1  
SOFTWARE: WORD PERFECT 6.1 and ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/099,753  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/603,753  
FILING DATE: 20 FEB 1996  
APPLICATION NUMBER: U.S. 08/373,799  
FILING DATE: 17 JAN 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: ARLES A. TAYLOR, JR.  
REGISTRATION NUMBER: 39,395  
REFERENCE/DOCKET NUMBER: 1242/2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 493-8000  
TELEFAX: (919) 419-0383  
TELEX:  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: no  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: Bovine  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
TISSUE TYPE:  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
FEATURE:  
NAME/KEY: Chromogranin B Granin Sequence  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:

TELEX: INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: no  
ANTI-SENSE: no  
ORIGINAL SOURCE: ORGANISM: Bovine  
INDIVIDUAL ISOLATE: DEVELOPMENTAL STAGE:  
TISSUE TYPE:  
CELL LINE:  
ORGANELLE:  
FEATURE: NAME/KEY: Chromogranin B Granin Sequence  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
PAGES:  
DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 18:  
US-08-603-753D-18

Query Match 100.0%; Score 18; DB 2; length 10;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4  
DB 2 NLAA 5

RESULT 4  
US-08-603-753D-19  
Sequence 19, Application US/08603753D  
Patent No. 5891857  
GENERAL INFORMATION:  
APPLICANT: HOLT, JEFFREY T.  
APPLICANT: JENSEN, ROY A.  
APPLICANT: PAGE, DAVID L.  
APPLICANT: KING, MARY-CLAIRE  
APPLICANT: SZABO, CSILLA I.  
APPLICANT: JETTTON, THOMAS L.  
APPLICANT: ROBINSON-BENION, CHERYL L.  
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2  
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON  
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARLES A. TAYLOR, JR.  
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER  
STREET: BOULEVARD  
CITY: DURHAM  
STATE: NORTH CAROLINA  
COUNTRY: USA  
ZIP: 27707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 KB storage  
COMPUTER: IBM PC/XT/AT compatible  
OPERATING SYSTEM: Windows 3.1  
SOFTWARE: WORD PERFECT 6.1 and ASCII

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/603,753D  
FILING DATE: 20 FEB 1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/373,799  
FILING DATE: 17 JAN 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: ARLES A. TAYLOR, JR.  
REGISTRATION NUMBER: 39,395  
REFERENCE/DOCKET NUMBER: 1242/2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 493-8000  
TELEFAX: (919) 419-0383  
TELEX:  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: no  
ANTI-SENSE: no  
ORIGINAL SOURCE: ORGANISM: Mouse  
INDIVIDUAL ISOLATE: DEVELOPMENTAL STAGE:  
TISSUE TYPE:  
CELL LINE:  
ORGANELLE:  
FEATURE: NAME/KEY: Chromogranin B Granin Sequence  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
PAGES:  
DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 19:  
US-08-603-753D-19

Query Match 100.0%; Score 18; DB 2; length 10;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4  
DB 2 NLAA 5

RESULT 5  
US-09-099-753-17  
Sequence 17, Application US/09099753  
Patent No. 6148903  
GENERAL INFORMATION:  
APPLICANT: HOLT, JEFFREY T.  
APPLICANT: JENSEN, ROY A.  
APPLICANT: PAGE, DAVID L.  
APPLICANT: KING, MARY-CLAIRE  
APPLICANT: SZABO, CSILLA I.  
APPLICANT: JETTTON, THOMAS L.  
APPLICANT: ROBINSON-BENION, CHERYL L.  
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2  
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON  
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.



AUTHORS: Masayoshi MISHINA  
TITLE: NOVEL PROTEINS AND GENES CODING THE SAME  
RELEVANT RESIDUES IN SEQ ID NO: 13: FROM 1 TO 7  
US-08-026-138E-13

Query Match 100.0%; Score 18; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4  
||||  
DB 4 NLAA 7

RESULT 2  
US-08-603-753D-17  
Sequence 17, Application US/08603753D  
Patent No. 5691857  
GENERAL INFORMATION:  
APPLICANT: HOLT, JEFFREY T.  
APPLICANT: JENSEN, ROY A.  
APPLICANT: PAGE, DAVID L.  
APPLICANT: KING, MARY-CLAIRE  
APPLICANT: SZABO, CSILLA I.  
APPLICANT: JETTTON, THOMAS L.  
APPLICANT: ROBINSON-BENTON, CHERYL L.  
APPLICANT: THOMPSON, MARILYN E.  
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2  
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON  
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARLES A. TAYLOR, JR.  
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER  
STREET: BOULEVARD  
CITY: DURHAM  
STATE: NORTH CAROLINA  
COUNTRY: USA  
ZIP: 27707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
COMPUTER: IBM PC/XT/AT compatible  
OPERATING SYSTEM: Windows 3.1  
SOFTWARE: WORD PERFECT 6.1 and ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/603,753D  
FILING DATE: 20 FEB 1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/373,799  
FILING DATE: 17 JAN 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: ARLES A. TAYLOR, JR.  
REGISTRATION NUMBER: 39,395  
REFERENCE/DOCKET NUMBER: 1242/2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 493-8000  
TELEFAX: (919) 419-0383  
TELEX:  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: no  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: Human  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:

TISSUE TYPE:  
CELL TYPE:  
CELL LINE:  
ORGANELLE:

FEATURE:  
NAME/KEY: Chromogranin B Granin Sequence  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS:

TITLE:  
JOURNAL:  
VOLUME:  
PAGES:  
DATE:

RELEVANT RESIDUES IN SEQ ID NO: 17:  
US-08-603-753D-17

Query Match 100.0%; Score 18; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. NO. 20;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4  
||||  
DB 2 NLAA 5

RESULT 3  
US-08-603-753D-18  
Sequence 18, Application US/08603753D  
Patent No. 5691857  
GENERAL INFORMATION:  
APPLICANT: HOLT, JEFFREY T.  
APPLICANT: JENSEN, ROY A.  
APPLICANT: PAGE, DAVID L.  
APPLICANT: KING, MARY-CLAIRE  
APPLICANT: SZABO, CSILLA I.  
APPLICANT: JETTTON, THOMAS L.  
APPLICANT: ROBINSON-BENTON, CHERYL L.  
APPLICANT: THOMPSON, MARILYN E.  
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2  
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON  
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARLES A. TAYLOR, JR.  
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER  
STREET: BOULEVARD  
CITY: DURHAM  
STATE: NORTH CAROLINA  
COUNTRY: USA  
ZIP: 27707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
COMPUTER: IBM PC/XT/AT compatible  
OPERATING SYSTEM: Windows 3.1  
SOFTWARE: WORD PERFECT 6.1 and ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/603,753D  
FILING DATE: 20 FEB 1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 08/373,799  
FILING DATE: 17 JAN 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: ARLES A. TAYLOR, JR.  
REGISTRATION NUMBER: 39,395  
REFERENCE/DOCKET NUMBER: 1242/2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 493-8000  
TELEFAX: (919) 419-0383

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:39 ; Search time 113.12 Seconds  
(without alignments)  
0.728 Million cell updates/sec

Title: US-09-603-713-29

Perfect score: 18

Sequence: 1 NLAA 4

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2-6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2-6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2-6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2-6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2-6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2-6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	7	1	US-08-026-138E-13
2	18	100.0	10	2	US-08-603-753P-17
3	18	100.0	10	2	US-08-603-753P-18
4	18	100.0	10	2	US-08-603-753P-19
5	18	100.0	10	4	US-09-099-753-17
6	18	100.0	10	4	US-09-099-753-18
7	18	100.0	10	4	US-09-099-753-19
8	18	100.0	10	4	US-08-986-106-14
9	18	100.0	10	4	US-08-986-106-15
10	18	100.0	10	4	US-08-986-106-16
11	18	100.0	12	1	US-08-258-639A-5
12	18	100.0	12	2	US-08-900-951-5
13	18	100.0	12	5	PCT-US95-07391A-5
14	18	100.0	14	1	US-07-961-837-4
15	18	100.0	14	1	US-08-232-453A-17
16	18	100.0	14	1	US-08-232-453A-25
17	18	100.0	17	3	US-09-101-146-52
18	18	100.0	29	2	US-08-332-562A-16
19	18	100.0	29	2	US-08-332-562A-20
20	18	100.0	33	1	US-08-669-791C-2
21	18	100.0	33	1	US-08-669-791C-5
22	18	100.0	33	2	US-08-669-790C-4
23	18	100.0	33	2	US-08-808-825-5
24	18	100.0	33	4	US-08-835-538C-2
25	18	100.0	41	1	US-07-956-700B-99
26	18	100.0	41	1	US-08-476-537-99
27	18	100.0	41	1	US-08-485-607-99

28	18	100.0	41	2	US-08-475-879-99	Sequence 99, Appl
29	18	100.0	43	1	US-08-377-687-37	Sequence 37, Appl
30	18	100.0	43	1	US-08-377-687-38	Sequence 38, Appl
31	18	100.0	43	2	US-08-777-192-37	Sequence 37, Appl
32	18	100.0	43	2	US-08-777-192-38	Sequence 38, Appl
33	18	100.0	43	4	US-08-971-982-37	Sequence 37, Appl
34	18	100.0	43	4	US-08-971-982-38	Sequence 38, Appl
35	18	100.0	48	1	US-08-062-472B-47	Sequence 47, Appl
36	18	100.0	57	1	US-08-358-160-94	Sequence 94, Appl
37	18	100.0	57	6	5466783-9	Patent No. 5466783
38	18	100.0	66	4	US-08-975-762-12	Sequence 12, Appl
39	18	100.0	66	4	US-08-821-324-12	Sequence 12, Appl
40	18	100.0	68	6	5198542-13	Patent No. 5198542
41	18	100.0	80	2	US-08-459-568-63	Sequence 63, Appl
42	18	100.0	80	2	US-08-399-411-63	Sequence 63, Appl
43	18	100.0	80	3	US-08-516-859A-63	Sequence 63, Appl
44	18	100.0	86	2	US-08-222-617A-11	Sequence 11, Appl
45	18	100.0	88	1	US-08-471-058-21	Sequence 21, Appl

#### ALIGNMENTS

RESULT 1  
US-08-026-138E-13  
Sequence 13, Application US/08026138E  
Patent No. 5502166  
GENERAL INFORMATION:  
APPLICANT: Masayoshi MISHINA  
TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nishiohata Residence 1-107  
STREET: 5214, Nishiohata-machi  
CITY: Niigata-shi  
STATE: Niigata-ken  
COUNTRY: JAPAN  
ZIP: 951  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
OPERATING SYSTEM: MS-DOS v.5  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/026,138E  
FILING DATE: 26-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 39563/1992  
FILING DATE: 26-FEB-1992  
APPLICATION NUMBER: JP 173155/1992  
FILING DATE: 30-JUN-1992  
APPLICATION NUMBER: JP 215017/1992  
FILING DATE: 12-AUG-1992  
APPLICATION NUMBER: JP 303878/1992  
FILING DATE: 13-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Hamburg, C. Bruce  
REGISTRATION NUMBER: 22,389  
REFERENCE/DOCKET NUMBER: F-4551  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 986-2340  
TELEFAX: (212) 953-7733  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single strand  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: mouse  
TISSUE TYPE: brain  
PUBLICATION INFORMATION:



Fri Sep 7 10:58:16 2001

us-09-603-713-29.rpt

---

R:Klein, J.M.; Mohrher, C.J.; Sleutels, F.; Riehm, J.P.; Rao, K.R.  
 Biochem. Biophys. Res. Commun. 205: 410-416, 1994  
 A:Title: Molecular cloning of two pigment-dispersing hormone (PDH) precursors in the blu  
 A:Reference number: JC2427; MUID:95091758  
 A:Accession: JC2427  
 A:Molecule type: mRNA  
 A:Residues: 1-78 <KLE>  
 A:Cross-references: GB:J36717; NID:9598090; PIDN:AAA67053.1; PID:9609439  
 A:Experimental source: eyestalk ganglia  
 C:Comment: This protein belongs to a family of structurally related neuropeptides.  
 C:Keywords: amidated carboxyl end  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-57/Domain: propeptide #status predicted <PRO>  
 F:58-75/Product: pigment-dispersing hormone #status predicted <MAT>  
 F:75/Modified site: amidated carboxyl end (Ala) (amide in mature form from following gly

Query Match 100.0%; Score 18; DB 2; Length 78;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 NLAA 4  
 ||||  
 Db 34 NLAA 37

## RESULT 13

T26945  
 hypothetical protein Y45F3A.5 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T26945  
 R:Steward, C.  
 submitted to the EMBL Data Library, October 1998  
 A:Reference number: Z20289  
 A:Accession: T26945  
 A:Status: preliminary; translated from GR/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-81 <MLT>  
 A:Cross-references: EMBL:AL032621; PIDN:CAA21493.1; GSPDB:GN00021; CESP:Y45F3A.5  
 A:Experimental source: clone Y45F3A  
 C:Genetics:  
 A:Gene: CESP:Y45F3A.5  
 A:Map position: 3  
 A:Introns: 27/3; 53/3

Query Match 100.0%; Score 18; DB 2; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 NLAA 4  
 ||||  
 Db 21 NLAA 24

## RESULT 14

W1BP27  
 gene 1.2 protein - phage T7  
 C:Species: phage T7  
 C:Date: 28-Feb-1981 #sequence\_revision 01-Sep-1981 #text\_change 23-Jul-1999  
 C:Accession: F43002; F43004; B93851; S42291; A04405  
 R:Dunn, J.J.; Thompson, K.  
 submitted to the Nucleic Acid Sequence Database, September 1982  
 A:Reference number: A94615  
 A:Accession: F43002  
 A:Molecule type: DNA  
 A:Residues: 1-85 <DUN>  
 R:Dunn, J.J.; Studier, F.W.  
 J. Mol. Biol. 148: 303-330, 1981  
 A:Title: Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the be  
 A:Reference number: A92866; MUID:82078034  
 A:Accession: F43004

A:Molecule type: DNA  
 A:Residues: 1-85 <DUN>  
 A:Cross-references: GB:V01127; NID:915498; PIDN:CAA24335.1; PID:915507  
 R:Saito, H.; Tabor, S.; Tamanoi, F.; Richardson, C.C.  
 Proc. Natl. Acad. Sci. U.S.A. 77: 3917-3921, 1980  
 A:Title: Nucleotide sequence of the primary origin of bacteriophage T7 DNA replicatio  
 A:Reference number: A93851; MUID:81054683  
 A:Accession: B93851

A:Molecule type: DNA  
 A:Residues: 1-85 <SAI>  
 A:Cross-references: GB:V01126; NID:915495; PIDN:CAA24325.1; PID:915497  
 R:Dunn, J.J.; Studier, F.W.  
 J. Mol. Biol. 166: 477-535, 1983  
 A:Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7  
 A:Reference number: S42283; MUID:83241725  
 A:Accession: S42291

A:Molecule type: DNA  
 A:Residues: 1-85 <DUN>  
 A:Cross-references: EMBL:V01146; NID:9431187; PIDN:CAA24392.1; PID:915571  
 A>Note: the authors did not translate the codon for residue 1  
 C:Comment: This early gene protein has been implicated in T7 DNA replication.  
 C:Genetics:  
 A:Gene: 1.2  
 A:Map position: 15.34-15.98  
 C:Superfamily: phage T7 gene 1.2 protein

Query Match 100.0%; Score 18; DB 1; Length 85;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 NLAA 4  
 ||||  
 Db 8 NLAA 11

## RESULT 15

T43965  
 hypothetical protein U6 [imported] - human herpesvirus 6 (strain HST) (fragment)  
 C:Species: human herpesvirus 6  
 A:Variety: strain HST  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
 C:Accession: T43965  
 R:Isegawa, Y.; Mukai, T.; Nakano, K.; Kagawa, M.; Chen, J.; Morl, Y.; Sunagawa, T.; K  
 J. Virol. 73: 8053-8063, 1999  
 A:Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A a  
 A:Reference number: Z22732; MUID:99412319  
 A:Accession: T43965  
 A:Status: preliminary; translated from GR/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-85 <ISE>  
 A:Cross-references: EMBL:AB021506; NID:94995977; PIDN:BAAT8226.1; PID:94995993  
 A:Experimental source: strain HST; pop. variant B  
 C:Genetics:  
 A>Note: U6

Query Match 100.0%; Score 18; DB 2; Length 85;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 NLAA 4  
 ||||  
 Db 20 NLAA 23

Search completed: September 6, 2001, 16:45:56  
 Job time: 500 sec

```

RESULT 8
G82647
hypothetical protein XF1694 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: G82647
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: G82647
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-62 <SIM>
A:Cross-references: GB:AE003994; GB:AE003849; NID:99106756; PIDN:MAF84503.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; From
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Matsumura, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1694

Query Match          100.0%; Score 18; DB 2; Length 62;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4
|||||
DB 22 NLAA 25

RESULT 9
S70550
ssPA protein - Salmonella typhimurium (fragment)
C:Species: Salmonella typhimurium
C:Date: 13-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 08-Oct-1999
C:Accession: S70550; S70455
R:Hueck, C.J.; Hantman, M.J.; Bajaj, V.; Johnston, C.; Lee, C.A.; Miller, S.L.
Mol. Microbiol. 18, 479-490, 1995
A:Title: Salmonella typhimurium secreted invasion determinants are homologous to Shigell
A:Reference number: S70455; MUID:96342378
A:Accession: S70550
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-64 <HND>
A:Cross-references: EMBL:U030491; NID:q1117811; PIDN:AAC43549.1; PID:q1117815
A:Experimental source: strain SL1344
A:Accession: S70455
A:Molecule type: protein
A:Residues: 2-13 <HND>
C:Genetics:
A:Gene: ssPA
A:Map position: 59

Query Match          100.0%; Score 18; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 NLAA 4
|||||
DB 26 NLAA 29

RESULT 10
T44154
hypothetical protein U6 [imported] - human herpesvirus 6 (strain Z29)
C:Species: human herpesvirus 6
A:Variety: strain Z29
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Jun-2000
C:Accession: T44154
R:Dominguez, G.; Dambaugh, T.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pellett, P.E.
J. Virol. 73, 8040-8052, 1999
A:Title: Human herpesvirus 6B genome sequence: coding content and comparison with hum
A:Reference number: Z22734; MUID:99412318
A:Accession: T44154
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-68 <DOM>
A:Cross-references: EMBL:AF157706; PIDN:AAD49624.1
A:Experimental source: strain Z29; variant B
C:Genetics:
A:Note: U6

Query Match          100.0%; Score 18; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4
|||||
DB 3 NLAA 6

RESULT 11
B83312
hypothetical protein PA2668 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83312
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; L
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: B83312
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-69 <STO>
A:Cross-references: GB:AE004695; GB:AE004091; NID:99948730; PIDN:AAG06056.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2668

Query Match          100.0%; Score 18; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4
|||||
DB 53 NLAA 56

RESULT 12
JC2427
pigment-dispersing hormone II precursor - blue crab
N:Alternate names: PDH2
C:Species: Callinectes sapidus (blue crab)
C:Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 21-Jul-2000
C:Accession: JC2427

```

```
RESULT 3
T36103
hypothetical protein SCE134.21 - Streptomyces coelicolor (fragment)
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36103
R:Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrett, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A:Reference number: 221596
A:Accession: T36103
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-31 <SAU>
A:Cross-references: EMBL:AL049661; PIDN:CAB41219.1; GSPDE:GN00070; SCOEDB:SCE134.21
A:Experimental source: strain A5(2)
C:Genetics:
A:Gene: SCOEDB:SCE134.21

Query Match
Best Local Similarity 100.0%; Score 18; DB 2; Length 31;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLA 4
DB 28 NLA 31

RESULT 4
A42598
urease accessory protein urec - Escherichia coli (fragment)
C:Species: Escherichia coli
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 19-May-2000
C:Accession: A42598
R:Collins, C.M.; Gutman, D.M.
J. Bacteriol. 174, 883-888, 1992
A:Title: Insertional inactivation of an Escherichia coli urease gene by IS3411.
A:Reference number: A42598; MUID:92121127
A:Accession: 1021
A:Accession: A42598
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-42 <COL>
A:Cross-references: GB:S77596; NID:g243731; PIDN:AB21161.1; PID:g243732
A:Note: Sequence extracted from NCBI backbone (NCBIN:77596, NCBI:P:77597)
C:Superfamily: hydrogenase expression/formation protein hypB

Query Match
Best Local Similarity 100.0%; Score 18; DB 2; Length 42;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLA 4
DB 14 NLA 17

RESULT 5
G83867
hypothetical protein BH1743 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: G83867
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20263114
A:Accession: G83867
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-44 <STO>
```

```
A:Cross-references: GB:AF001513; GB:BA000004; NID:g10174345; PIDN:BA05462.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1743

Query Match
Best Local Similarity 100.0%; Score 18; DB 2; Length 44;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLA 4
DB 15 NLA 18

RESULT 6
W2BP22
arf protein - phage P22
C:Species: phage P22
A:Note: host Salmonella typhimurium
C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 23-Jul-1999
C:Accession: S04249
R:Semerjian, A.V.; Malloy, D.C.; Potete, A.R.
J. Mol. Biol. 207, 1-13, 1989
A:Title: Genetic structure of the bacteriophage P22 P(L) operon.
A:Reference number: S04245; MUID:89293845
A:Accession: S04249
A:Molecule type: DNA
A:Residues: 1-47 <SEM>
A:Cross-references: GB:X15637; NID:g15646; PIDN:CA33652.1; PID:g15653
C:Genetics:
A:Gene: arf
C:Superfamily: phage P22 arf protein

Query Match
Best Local Similarity 100.0%; Score 18; DB 1; Length 47;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLA 4
DB 27 NLA 30

RESULT 7
TTRIV2
venom basic proteinase inhibitor II - ringhals
C:Species: Hemachatus haemachatus, Sepeodon haemachatus (Ringhals)
C:Date: 30-Nov-1980 #sequence_revision 30-Nov-1980 #text_change 16-Aug-1996
C:Accession: A01216
R:Hokama, Y.; Iwaga, S.; Tatsu, T.; Suzuki, T.
J. Biochem. 79, 559-578, 1976
A:Title: Snake venom proteinase inhibitors. III. Isolation of five polypeptide inhibi
o acid sequences of two of them.
A:Reference number: A91942; MUID:76237547
A:Accession: A01216
A:Molecule type: protein
A:Residues: 1-57 <HOK>
A:Comment: The activity of this inhibitor is similar to that of bovine basic protease
C:Superfamily: basic proteinase inhibitor; animal Kunitz-type proteinase inhibitor ho
C:Keywords: serine proteinase inhibitor; venom
F:5-55/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
F:5-55,14-38,30-51/Disulfide bonds: #status predicted

Query Match
Best Local Similarity 100.0%; Score 18; DB 1; Length 57;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLA 4
DB 24 NLA 27
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:55 ; Search time 134.15 Seconds  
(without alignments)  
2.271 Million cell updates/sec

Title: US-09-603-713-29

Perfect score: 18  
Sequence: 1 NLAA 4

Scoring table: BLAST62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

PIR-68:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	18 100.0	14 2	B39111	Ig heavy chain V r
2	18 100.0	30 2	A44913	34k core flagella
3	18 100.0	31 2	T36103	hypothetical prote
4	18 100.0	42 2	A42598	urase accessory p
5	18 100.0	44 2	G83867	hypothetical prote
6	18 100.0	47 1	W2BP22	art protein - phag
7	18 100.0	57 1	TIRIV2	venom basic protei
8	18 100.0	62 2	G82647	hypothetical prote
9	18 100.0	64 2	S70550	sppA protein - sal
10	18 100.0	68 2	T44154	hypothetical prote
11	18 100.0	69 2	B83112	hypothetical prote
12	18 100.0	78 2	JC2427	hypothetical prote
13	18 100.0	81 2	T26945	pigment-dispersin
14	18 100.0	85 1	W1BP27	hypothetical prote
15	18 100.0	85 2	T43965	gene 1.2 protein
16	18 100.0	90 2	S77887	hypothetical prote
17	18 100.0	91 2	H69543	DNA-binding protei
18	18 100.0	92 2	A28762	NADH dehydrogenase
19	18 100.0	102 2	D72052	Ig kappa chain V r
20	18 100.0	102 2	H86572	thioredoxin (import
21	18 100.0	106 2	T36973	thioredoxin (import
22	18 100.0	107 2	C82100	cytochrome c554 yc
23	18 100.0	108 2	J00856	hypothetical 12.2k
24	18 100.0	110 2	B64480	hypothetical prote
25	18 100.0	111 1	A52233	thioredoxin 2 (val
26	18 100.0	111 1	S07726	hypothetical prote
27	18 100.0	117 2	T14465	lipid transfer pro
28	18 100.0	117 2	T07861	germination-specif
29	18 100.0	117 2	T07866	germination-specif

30	18 100.0	117 2	E86458	RNA-binding protei
31	18 100.0	118 2	C84806	probable nonspecif
32	18 100.0	118 2	T31183	hypothetical prote
33	18 100.0	120 2	T39623	probable adenylate
34	18 100.0	121 1	A36270	hemoglobin - Tetra
35	18 100.0	121 1	S32556	hemoglobin - Tetra
36	18 100.0	123 2	T33456	hypothetical prote
37	18 100.0	124 2	A45691	hypothetical prote
38	18 100.0	126 2	T10191	hypothetical prote
39	18 100.0	127 2	H82442	hypothetical prote
40	18 100.0	127 2	A96509	protein F27F5.9 (1
41	18 100.0	128 2	H81919	probable glycine c
42	18 100.0	128 2	S81183	glycine cleavage c
43	18 100.0	128 2	D75465	hypothetical prote
44	18 100.0	128 2	S27917	hypothetical prote
45	18 100.0	128 2	JQ2180	hypothetical 14.8k

## ALIGNMENTS

RESULT 1  
B39111  
Ig heavy chain V region - Pacific hagfish (fragment)  
C:Species: Eptatretus stoutii (Pacific hagfish)  
C:Date: 18-Oct-1991 #sequence\_revision 18-Oct-1991 #text\_change 16-Aug-1996  
C:Accession: B39111  
R:Varner, J.; Neame, P.; Litman, G.W.  
Proc. Natl. Acad. Sci. U.S.A. 88, 1746-1750, 1991  
A:Title: A serum heterodimer from hagfish (Eptatretus stoutii) exhibits structural sl  
A:Reference number: A39111; MUID:91156684  
A:Accession: B39111  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-14 <VAR>  
C:Keywords: heterotetramer; immunoglobulin

Query Match 100.0%; Score 18; DB 2; Length 14;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4  
DB 6 NLAA 9

RESULT 2  
A44913  
34k core flagella protein - Leptospira interrogans (fragment)  
C:Species: Leptospira interrogans  
C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 23-Mar-1995  
C:Accession: A44913; A11210  
R:Itunaba, G.A.; Bolin, C.A.; Zuercher, R.L.  
J. Bacteriol. 174, 4761-4768, 1992  
A:Title: Characterization of the periplasmic flagellum proteins of Leptospira interro  
A:Reference number: A44913; MUID:92325069  
A:Accession: A44913  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-30 <TRU>  
A:Experimental source: sv. pomona type kennewicki  
C:Superfamily: flagellin

Query Match 100.0%; Score 18; DB 2; Length 30;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4  
DB 6 NLAA 9





DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE NONSPECIFIC LIPID-TRANSFER PROTEIN 1 PRECURSOR (LTP 1).  
 GN LTP1.  
 OS Brassica napus (Rape).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. DARMOR;  
 RX MEDLINE=96275957; PubMed=8680310;  
 RA Soulier I.A., Vergnolle C., Miginiac E., Kader J.-C.;  
 RT "Germination-specific lipid transfer protein cDNAs in Brassica napus  
 L.";  
 RL Planta 199:229-237(1996).  
 CC -1- FUNCTION: PLANT NONSPECIFIC LIPID-TRANSFER PROTEINS TRANSFER  
 CC PHOSPHOLIPIDS AS WELL AS GALACTOLIPIDS ACROSS MEMBRANES. MAY PLAY  
 CC A ROLE IN WAX OR CUTIN DEPOSITION IN THE CELL WALLS OF EXPANDING  
 CC EPIDERMAL CELLS AND CERTAIN SECRETORY TISSUES.  
 CC -1- SIMILARITY: BELONGS TO THE PLANT LTP FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U22105; AAB37228.1; -  
 CC DR HSSP: P19656; IAFH.  
 CC DR InterPro: IPR000528; -  
 CC DR Pfam: PF00279; LTP; 1.  
 CC DR PRINTS: PR00382; LIPIDTRANSFER.  
 CC DR PROSITE: PS00597; PLANT\_LTP; 1.  
 CC DR  
 CC KW Lipid-binding; Transport; Signal; Multigene family.  
 CC FT SIGNAL 1 25 POTENTIAL.  
 CC FT CHAIN 26 117 NONSPECIFIC LIPID-TRANSFER PROTEIN 1.  
 CC FT DISULFID 29 76 POTENTIAL.  
 CC FT DISULFID 39 53 POTENTIAL.  
 CC FT DISULFID 54 99 POTENTIAL.  
 CC FT DISULFID 74 113 POTENTIAL.  
 CC SQ SEQUENCE 117 AA: 11857 MW: 44873C5970509AE1 CRC64;

Query Match 100.0%; Score 18; DB 1; Length 117;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 NLAA 4  
 Db 35 NLAA 38

Search completed: September 6, 2001, 16:51:12  
 Job time: 816 sec

```

CC      H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
CC      -1- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: Y16633; CAA76338.1; -.
CC      InterPro: IPR000164; -.
CC      InterPro: IPR000166; -.
CC      Pfam: PF00125; histone; 1.
CC      PROSITE: PS00322; HISTONE_H3_1; FALSE_NEG.
CC      PROSITE: PS00359; HISTONE_H3_2; 1.
CC      Nucleic protein; Chromosomal protein; DNA-binding; Nucleosome core.
CC      FT NON_TER 1 114 114
CC      SEQUENCE 114 AA; 12689 MW; 36A9B6EA7F1D01FC CRC64;

Query Match
Best Local Similarity 100.0%; Score 18; DB 1; Length 114;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 NLAA 4
DB      101 NLAA 104

RESULT 13
H39_STYLE
ID      H39_STYLE STANDARD; PRT; 114 AA.
AC      P81196;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      HISTONE H3-2 (FRAGMENT).
CN      H3-2.
OS      Stylonychia lemnae.
OC      Eukaryota; Alveolata; Ciliophora; hypotrichs; stichotrichida;
OC      Oxytrichidae; Stylonychia.
OX      NCBI_TaxID=5949;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Bernhard D.;
RT      Several highly divergent histone H3 genes in the hypotrich ciliate
RT      Stylonychia lemnae.
RL      Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
CC      -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE
CC      IN NUCLEOSOME FORMATION.
CC      -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC      H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATIVELY 146 BP OF DNA.
CC      -1- SIMILARITY: BELONGS TO THE HISTONE H3 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: Y16635; CAA76338.1; -.
CC      InterPro: IPR000164; -.
CC      InterPro: IPR000166; -.
CC      Pfam: PF00125; histone; 1.
CC      PROSITE: PS00322; HISTONE_H3_1; FALSE_NEG.
CC      PROSITE: PS00359; HISTONE_H3_2; 1.
CC      Nucleic protein; Chromosomal protein; DNA-binding; Nucleosome core.
CC      FT NON_TER 1 114 114
CC      SEQUENCE 114 AA; 12689 MW; 36A9B6EA7F1D01FC CRC64;

```

```

FT      NON_TER 114 114
SQ      SEQUENCE 114 AA; 12689 MW; 36A9B6EA7F1D01FC CRC64;

Query Match
Best Local Similarity 100.0%; Score 18; DB 1; Length 114;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 NLAA 4
DB      101 NLAA 104

RESULT 14
DOC1_HUMAN
ID      DOC1_HUMAN STANDARD; PRT; 115 AA.
AC      O14519;
DT      15-JUL-1999 (Rel. 38, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      PUTATIVE ORAL CANCER SUPPRESSOR (DELETED IN ORAL CANCER-1) (DOC-1).
CN      DOC1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE-97472613; PubMed-9331572;
RA      Dalgo Y., Suzuki K., Maruyama O., Miyoshi Y., Yasuda T., Kabuto T.,
RA      Imaoka S., Fujiwara T., Takahashi E., Fujino M.A., Nakamura Y.;
RT      *Isolation, mapping and mutation analysis of a human cDNA homologous
RT      to the doc-1 gene of the Chinese hamster, a candidate tumor
RT      suppressor for oral cancer.
RL      Genes Chromosomes Cancer 20:204-207(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE-Testis;
RA      Duh F.-M., Latif F., Wong D., Popescu N., Lerman M.I.;
RT      *Cloning and characterization of the human homolog of the hamster oral
RT      tumor suppressor gene doc-1.
RL      Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
CC      -1- SIMILARITY: BELONGS TO THE DOC1 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AB006077; BAA22937.1; -.
CC      EMBL: AF006484; AAC77831.1; -.
CC      MIM: 602198; -.
CC      Anti-oncogene.
CC      SEQUENCE 115 AA; 12365 MW; F3149F72467598C2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 18; DB 1; Length 115;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 NLAA 4
DB      6 NLAA 9

RESULT 15
NTL1_BRANA
ID      NTL1_BRANA STANDARD; PRT; 117 AA.
AC      Q42614;
DT      01-NOV-1997 (Rel. 35, Created)

```

SQ SEQUENCE 110 AA; 12235 MW; OCFBD6812B8C888 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 18; DB 1; Length 110;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
1111  
DB 43 NLAA 46

RESULT 10  
ID YE43\_METJA STANDARD; PRT; 110 AA.  
AC 058838;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL PROTEIN MJ1443.  
GN MJ1443.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
OC Methanococcus.  
OC NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake O., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Meldrum J.F., Fuhmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE UPF0132 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
DR EMBL; U67585; AAB99453.1; -.  
DR TIGR; MJ1443; -.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 15 35 POTENTIAL.  
FT TRANSMEM 49 69 POTENTIAL.  
FT TRANSMEM 70 90 POTENTIAL.  
FT SEQUENCE 110 AA; 12411 MW; OD16244307741796 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 18; DB 1; Length 110;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
1111  
DB 76 NLAA 79

RESULT 11  
ID YM01\_PARTE STANDARD; PRT; 113 AA.  
AC P15603;

DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-APR-1990 (Rel. 14, Last annotation update)  
DE HYPOTHETICAL 12.8 KDA PROTEIN (ORF1).  
OS Paramacium tetraurelia.  
OC Mitochondrion.  
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Pericollida;  
OC Paramacium.  
OC NCBI\_TaxID=5688;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=STOCK 51;  
RX MEDLINE=90174913; PubMed=2308823;  
RA Pritchard A.E., Sellhammer J.J., Mahalingam R., Sable C.L.,  
RA Venuti S.E., Cummings D.J.;  
RT "Nucleotide sequence of the mitochondrial genome of Paramacium.";  
RL Nucleic Acids Res. 18:173-180(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=STOCK 51;  
RX MEDLINE=87055241; PubMed=3023187;  
RA Pritchard A.E., Sellhammer J.J., Cummings D.J.;  
RT "Paramacium mitochondrial DNA sequences and RNA transcripts for  
RT cytochrome oxidase subunit I, URF1, and three ORFs adjacent to the  
RT replication origin.";  
RL Gene 44:243-253(1986).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
DR EMBL; X15917; CAA34059.1; -.  
DR EMBL; M15275; AAA79261.1; -.  
DR PIR; S07726; S07726.  
KW Hypothetical protein; Mitochondrion.  
KW SEQUENCE 113 AA; 12857 MW; EE1EE299DAE849E3 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 18; DB 1; Length 113;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
1111  
DB 89 NLAA 92

RESULT 12  
ID H37\_STYLE STANDARD; PRT; 114 AA.  
AC P81201;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HISTONE H3-7 (FRAGMENT).  
GN H3-7.  
OS Stylyonychia lemnae.  
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichida;  
OC Oxytrichidae; Stylyonychia.  
OC NCBI\_TaxID=5949;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Bernhard D.;  
RT "Several highly divergent histone H3 genes in the hypotrich ciliate  
RT Stylyonychia lemnae.";  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HISTONE H3, ALONG WITH HISTONE H4, PLAYS A CENTRAL ROLE  
CC IN NUCLEOSOME FORMATION.  
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF

```

RA Read T.D., Brinham R.C., Shen C., Gill S.R., Haidelberg J.F.,
RA White O., Hickey E.R., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson S.W., Deboy R., Kolony J., McClarty G., Salberg S.L.,
RA Eilen J., Fraser C.M.,
RT "Genome sequences of Chlamydia trachomatis MOpn and Chlamydia
RT pneumoniae AR39."
RT Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE-20330349; PubMed-10871362;
RA Shiral M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishi K., Hattori M., Kohara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RL from Japan and CML029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC TO A DISULFIDE, AND CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.sdb.ch).
CC -----
CC EMBL: AE001648; AAD18798.1; -
CC EMBL: AE002172; AAF37973.1; -
CC EMBL: AP002547; BAA38866.1; -
CC TIGR: CP0088; -
CC HSSP: P00274; 1XOA.
DR InterPro: IPR000063; -
DR Pfam: PF00085; Unfolded; 1.
DR PRINTS: PR00421; THIOREDOXIN.
DR PROSITE: PS00194; THIOREDOXIN; 1.
KW Redox-active center; Electron transport.
FT DISULFID 28
FT SEQUENCE 102 AA; 11292 MW; 76A19021832ABA68 CMC64;
SO SEQUENCE

Query Match 100.0%; Score 18; DB 1; Length 102;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLA4 4
DB 40 NLA4 43

RESULT 8
AC FR7A_ECOLI STANDARD; PRT: 108 AA.
ID P21315;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE HYDROTHERMAL 12.2 KDA PROTEIN (ORFA) (RETRON EC67).
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
CC NCBI_TaxID=562;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CL-1;
RX MEDLINE-91067724; PubMed-1701261;
RA Hsu M.-Y., Inouye M., Inouye S.;
RT "Retron for the 67-base multicity single-stranded DNA from
RT Escherichia coli: a potential transposable element encoding both
RT reverse transcriptase and Dam methylase functions."

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 87:9454-9458(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; M55249; AAA2391.1; -.
DR PIR; J00856; J00856.
SQ TRANSDUCIBLE ELEMENT; HYPOTHETICAL PROTEIN.
SQ SEQUENCE 108 AA; 12193 MW; 050B940F2A20551 CRC64;
QY 1 NLA 4
Db 91 NLA 94
Query Match 100.0%; Score 18; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NLA 4
Db 91 NLA 94
RESULT 9
TH12_ANASP ID TH12_ANASP STANDARD; PRT; 110 AA.
AC P20857.
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE THIOREDOXIN 2 (TRX-2).
FRXB.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=89123014; PubMed=2492494;
RA Alam J, Curtis S.E., Gleason F.K., Gerami-Nejad M., Fuchs J.A.:
RT "Isolation, sequence, and expression in Escherichia coli of an
RT unusual chlorodioxin gene from the cyanobacterium Anabaena sp. strain
RT PCC 7120.";
RL J. Bacteriol. 171:162-171(1989).
[2]
RN 12
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE=96131880; PubMed=8590004;
RA Saarinen M., Gleason F.K., Eklund H.:
RT "Crystal structure of chlorodioxin-2 from Anabaena.";
RL Structure 3:1097-1108(1995).
CC -!- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; M22997; AAA22048.1; -.
DR PIR; A32233; A32233.
DR PDB; 1THX; 1S-OCT-95.
DR InterPro; IPR000063; -.
DR Pfam; PF00085; ThioRed; 1.
DR PRINTS; PR00421; THIOREDOXIN.
DR PROSITE; PS00194; THIOREDOXIN; 1.
RW Redox-active center; Electron transport; 3D-structure.
FT INT_MET 0
FT DISULFID 32 35 REDOX-ACTIVE (BY SIMILARITY).

```

OY 1 NLAA 4  
 DB 8 NLAA 11

RESULT 5  
 ID C552\_MARHY STANDARD: PRT: 88 AA.  
 AC P82903:  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DE CYTOCHROME C-552 (C552).  
 OS Marinobacter hydrocarbonoclasticus (Pseudomonas nautica).  
 OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;  
 OC Marinobacter group; Marinobacter.  
 NCBI\_TaxID=2743;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=617;  
 RA Sariva L.M., Fauque G., Besson S., Moura I.;  
 RT "Physical-chemical and spectroscopic properties of the monohemic  
 cytochrome C552 from Pseudomonas nautica 617.";  
 RL Eur. J. Biochem. 224:1011-1017(1994).  
 RN [2]  
 RP SEQUENCE.  
 RC STRAIN=617;  
 RA Maras B., Shina E., Di Leonardo L., Pitarì G., Malatesta F.;  
 RT Submitted (NOV-2000) to the SWISS-PROT data bank.  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS), AND MASS SPECTROMETRY.  
 RC STRAIN=617;  
 RA MEDLINE=92939395; PubMed=10369779;  
 RA Brown K., Nuriel D., Besson S., Shepard W., Moura J., Moura I.,  
 RA Pegoni M., Cambillau C.;  
 RT "MAD structure of Pseudomonas nautica dimeric cytochrome c552 mimicks  
 the c4 dihemetic cytochrome domain association.";  
 RL J. Mol. Biol. 289:1017-1028(1999).  
 CC -1- FUNCTION: ELECTRON DONOR FOR PERIPLASMIC NITRATE REDUCTASE.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.  
 CC -1- PTM: BINDS ONE HEME GROUP PER MOLECULE.  
 CC -1- MASS SPECTROMETRY: MW=9466; METHOD=MALDI.  
 CC -1- SIMILARITY: BELONGS TO THE CLASS I CYTOCHROME C FAMILY.  
 DR PDB: 1CNO; 22-JUL-99.  
 DR InterPro: IPR000345;  
 DR InterPro: IPR002329;  
 DR InterPro: IPR003088;  
 DR Pfam: PF00034; cytochrome\_c; 1.  
 DR Prosite: PS00190; CYTOCHROME\_C; 1.  
 KW Electron transport; Heme; Periplasmic; 3D-structure.  
 FT BINDING 14  
 FT BINDING 17  
 FT METAL 18  
 FT METAL 60  
 FT METAL 79  
 FT METAL 88  
 FT CONFLICT 79  
 FT CONFLICT 88  
 FT SEQUENCE 88 AA; 8832 MW; B98FEFB6C4ABZ5 CRC64;

Query Match 100.0%; Score 18; DB 1; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 DB 74 NLAA 77

RESULT 6  
 MENB-STILA  
 ID MENB-STILA STANDARD: PRT: 100 AA.  
 AC 024356;

DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE MEN-8 PROTEIN PRECURSOR.  
 GN MEN-8.  
 OS Silene latifolia.  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;  
 OC Caryophyllales; Caryophyllaceae; Silene.  
 NCBI\_TaxID=37657;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Flower;  
 RX MEDLINE=9377129; PubMed=9232878;  
 RA Scott C.P., Li Y., Robertson S.E., Willis M.E., Gilmartin P.M.;  
 RT "Sex determination in dioecious Silene latifolia. Effects of the Y  
 chromosome and the parasitic smut fungus (Ustilago violacea) on gene  
 expression during flower development.";  
 RL Plant Physiol. 114:969-979(1997).  
 CC -1- SIMILARITY: BELONGS TO THE A9 / FILL FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Y08780; CAA70033.1; .  
 DR  
 KM Signal. 1  
 FT CHAIN 100  
 FT DISULFID 38  
 FT DISULFID 76  
 FT DISULFID 48  
 FT DISULFID 65  
 FT DISULFID 66  
 FT DISULFID 91  
 FT DISULFID 98  
 FT SEQUENCE 100 AA; 10221 MW; 1E5AED654D319E36 CRC64;

Query Match 100.0%; Score 18; DB 1; Length 100;  
 Best Local Similarity 100.0%; Pred. No. 1;le+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 DB 92 NLAA 95

RESULT 7  
 THIO\_CHLPN STANDARD: PRT: 102 AA.  
 ID THIO\_CHLPN  
 AC Q927P5; Q9J0G1;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE THIOREDOXIN (TRX).  
 GN TRXA OR CPN0659 OR CP0088.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Chlamydia pneumoniae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CWL029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RN Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;

OK NCBI\_TaxID=10754;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89293845; PubMed=2738922;  
 RA Semerjian A.V., Malloy D.C., Potete A.R.;  
 RT "Genetic structure of the bacteriophage P22 PL operon."  
 RL J. Mol. Biol. 207:1-13(1989).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RA Kropinski A.M.B., Vanderbyl C.S.;  
 RT "The completed sequence of genome of Salmonella phage P22."  
 RL Submitted (DEC-1989) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: ARP MAY BE RECOMBINATION-RELATED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X15637; CAA3652.1; -  
 DR EMBL: AF217253; AAF75015.1; -  
 DR PIR: S04249; W2BP22;  
 SO SEQUENCE 47 AA: 5517 MW: 2C8682C7B4610FE3 CRC64;

Query Match 100.0%; Score 18; DB 1; Length 47;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 DB 27 NLAA 30

RESULT 3  
 ID 1VB2\_HEMHA STANDARD; PRT; 57 AA.  
 AC P00985;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-MAY-1992 (Rel. 22, Last annotation update)  
 DE VEMOM BASIC PROTEASE INHIBITOR II.  
 OS Hemachatus haemachatus (Ringhals) (Sespedon haemachatus).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidea;  
 CC Elapidae; Elapinae; Hemachatus.  
 CC NCB1\_TaxID=8626;  
 OX NCB1  
 RN (1)  
 RP SEQUENCE.  
 RX MEDLINE=76237547; PubMed=950337;  
 RA Hokama Y., Iwanaga S., Tatsuki T., Suzuki T.;  
 RT "Snake venom proteinase inhibitors. III. Isolation of five  
 RT polypeptide inhibitors from the venoms of Hemachatus haemachatus  
 RT (Ringhals's cobra) and Naja naja (Cape cobra) and the complete amino  
 RT acid sequences of two of them."  
 RL J. Biochem. 79:559-578(1976).  
 CC -1- FUNCTION: THE ACTIVITY OF THIS INHIBITOR IS SIMILAR TO THAT OF  
 CC BOVINE BASIC PROTEASE INHIBITOR.  
 CC -----  
 CC -1- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.  
 DR PIR: A01216; TIRIV2.  
 DR HSSP: P00981; IDTK.  
 DR InterPro: IPR002223; -  
 DR Pfam: PF00014; Kunitz\_BPTI; 1.  
 DR PRINTS: PRO0759; BASICPTASE.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE: PS50279; BPTI\_KUNITZ\_2; 1.  
 KW Venom; Serine protease inhibitor.  
 FT DISULFID 5 55 BY SIMILARITY.  
 FT DISULFID 14 38 BY SIMILARITY.  
 FT DISULFID 30 51 BY SIMILARITY.

FT ACT\_SITE 15 16 REACTIVE BOND (BY SIMILARITY).  
 SQ SEQUENCE 57 AA: 6407 MW: 37CF03D3A03D7F2A CRC64;

Query Match 100.0%; Score 18; DB 1; Length 57;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 DB 24 NLAA 27

RESULT 4  
 ID V12\_BPT7 STANDARD; PRT; 85 AA.  
 AC P03780;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE GENE 1.2 PROTEIN.  
 GN 1.2  
 OS Bacteriophage T7  
 CC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae;  
 CC T7-like phages.  
 CC NCB1\_TaxID=10760;  
 OX NCB1  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=83241725; PubMed=6864790;  
 RA Dunn J.J., Studier F.W.;  
 RT "Complete nucleotide sequence of bacteriophage T7 DNA and the  
 RT locations of T7 genetic elements."  
 RL J. Mol. Biol. 166:477-535(1983).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=82078034; PubMed=7310871;  
 RA Dunn J.J., Studier F.W.;  
 RT "Nucleotide sequence from the genetic left end of bacteriophage T7  
 RT DNA to the beginning of gene 4."  
 RL J. Mol. Biol. 148:303-330(1981).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=81054683; PubMed=6254001;  
 RA Saito H., Tabor S., Tamaol F., Richardson C.C.;  
 RT "Nucleotide sequence of the primary origin of bacteriophage T7 DNA  
 RT replication: relationship to adjacent genes and regulatory  
 RT elements."  
 RL Proc. Natl. Acad. Sci. U.S.A. 77:3917-3921(1980).  
 CC -1- FUNCTION: THIS IS AN INHIBITOR OF THE HOST (E.COLI) DGTP  
 CC TRIPHOSPHOHYDROLASE (DGTASE). IT IS IMPLICATED IN DNA  
 CC REPLICATION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: V01146; CAA24392.1; -  
 DR EMBL: V01126; CAA24325.1; -  
 DR EMBL: V01127; CAA24335.1; -  
 DR PIR: A04405; W1BP27.  
 DR PIR: S42291; S42291.  
 KW Early protein; DNA replication.  
 SO SEQUENCE 85 AA: 10190 MW: EE292B9102BE2650 CRC64;

Query Match 100.0%; Score 18; DB 1; Length 85;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:11 : Search time 72.75 Seconds  
(without alignments)  
1.883 Million cell updates/sec

Title: US-09-603-713-29

Perfect score: 18

Sequence: 1 NLAA 4

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 93435 segs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	18	100.0	43 1 NLTP_RAPSA	P29420 raphanus sa
2	18	100.0	47 1 VARF_BPP22	P14112 bacterioph
3	18	100.0	57 1 IVB2_HEMHA	P00985 hemachatus
4	18	100.0	85 1 VI2_BPT7	P03780 bacterioph
5	18	100.0	88 1 C552_MARBY	P82903 marinobacte
6	18	100.0	100 1 MEN8_STILA	G24356 silene latif
7	18	100.0	102 1 THIO_CULRN	O92735 chlamydia p
8	18	100.0	108 1 YR7A_ECOLI	P21315 escherichia
9	18	100.0	110 1 TH12_ANASP	P20857 anaplasma sp
10	18	100.0	110 1 YE43_METUA	G28838 methanococo
11	18	100.0	113 1 YM01_PARTE	P15603 paramedlum
12	18	100.0	114 1 H37_STYLE	P81196 stylyonchia
13	18	100.0	114 1 H39_STYLE	P81196 stylyonchia
14	18	100.0	115 1 DOC1_HUMAN	O14519 homo sapien
15	18	100.0	117 1 NLUT1_BRANA	O42614 brassica na
16	18	100.0	117 1 NLUT3_BRANA	O42614 brassica na
17	18	100.0	117 1 NLTR_BRAOL	O42642 brassica ol
18	18	100.0	118 1 NLTR_ARATH	O42580 arabidopsis
19	18	100.0	121 1 GLB_TERTY	P17724 tetrahymena
20	18	100.0	121 1 GLB_TERTH	O03459 tetrahymena
21	18	100.0	124 1 VSP1_BPLH	O04765 lactococcus
22	18	100.0	128 1 GCSH_NEIMA	O91761 neisseria m
23	18	100.0	128 1 GCSH_NEIMB	O91761 neisseria m
24	18	100.0	132 1 HSLU_SALTY	O30911 salmonella
25	18	100.0	132 1 CYCP_PARSP	P00143 paracoccus
26	18	100.0	132 1 YDDE_SALTY	P40788 salmonella
27	18	100.0	133 1 AAKG_PIG	O09138 sus scrofa
28	18	100.0	135 1 CY2_RHORO	P00092 rhodospiril
29	18	100.0	135 1 H3_YEAST	P02303 saccharomyc
30	18	100.0	143 1 YL6C_MYCLE	O69551 mycobacteri
31	18	100.0	143 1 YL6C_MYCTU	O69551 mycobacteri
32	18	100.0	144 1 HOLC_HAEIN	P43749 haemophilus
33	18	100.0	147 1 PHNB_ECOLI	P16681 escherichia

34	18	100.0	149 1 RL9_HAEIN	P44349 haemophilus
35	18	100.0	151 1 DUT_HAEIN	P43792 haemophilus
36	18	100.0	152 1 HSPA_BRAJA	P70917 bradyrhizob
37	18	100.0	152 1 RS11_SCHPO	P79013 schizosacch
38	18	100.0	155 1 RISC_AERPE	O9ydc5 aeropyrum p
39	18	100.0	155 1 RS7_THERH	P17291 thermus aqu
40	18	100.0	163 1 YHY0_YEAST	P38868 saccharomyc
41	18	100.0	170 1 RL17_MYCLE	O9x797 mycobacteri
42	18	100.0	173 1 SP12_YEAST	P17123 saccharomyc
43	18	100.0	174 1 CHCB_BOMMO	P08830 bombyx mori
44	18	100.0	174 1 PRIC_ECOLI	P23863 escherichia
45	18	100.0	175 1 LP20_HELPY	P53436 helicobacte

## ALIGNMENTS

RESULT 1  
ID NLTP\_RAPSA STANDARD: PRT: 43 AA.  
AC P29420:  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE SSED NONSPECIFIC LIPID TRANSFER PROTEIN-LIKE (FRAGMENT).  
OS Raphanus sativus (Radish).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicot; Rosidae; eurosid II;  
OC Brassicales; Brassicaceae; Raphanus.  
OX NCBI\_TaxID=3726;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Seed;  
RA Terms F.R.G., Goderis I.J., van Leuven F., Vanderleyden J.,  
RA Cammue B.P.A., Broekaert W.F.;  
RT "In vitro antifungal activity of a radish (Raphanus sativus L.) seed  
protein homologous to non-specific lipid transfer proteins.";  
RL Plant Physiol. 100:1055-1058(1992).  
CC - FUNCTION: PLANT-NONSPECIFIC LIPID-TRANSFER PROTEINS TRANSFER  
PHOSPHOLIPIDS AS WELL AS GALACTOLIPIDS ACROSS MEMBRANES (BY  
SIMILARITY).  
CC - FUNCTION: INHIBITS THE HYPHAL GROWTH OF SEVERAL FUNGI IN VITRO.  
CC - SUBUNIT: HOMODIMER.  
CC - SIMILARITY: BELONGS TO THE PLANT LTP FAMILY.  
DR HSP: P19656; IAPH.  
DR InterPro: IPR000528; .  
DR Pfam: PF00279; LTP; 1.  
DR PROSITE: PS00597; PLANT\_LTP; PARTIAL.  
KW Lipid-binding; Transport; Fungicide.  
FT NON\_TER 43  
SQ SEQUENCE 43 AA: 4311 MW: A723ABE610EDDC4F CRC64;

Query Match 100.0%; Score 18; DB 1; Length 43;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
DB 10 NLAA 13  
RESULT 2  
ID VARF\_BPP22 STANDARD: PRT: 47 AA.  
AC P14112:  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ACCESSORY RECOMBINATION FUNCTION PROTEIN.  
GN ARF.  
OS Bacteriophage P22.  
OC Bacteriophage P22.  
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.





RL Mol. Microbiol. 18:479-490(1995).  
 DR EMBL: U30491; AAC43549.1; -  
 FT NON\_TER 65  
 SQ SEQUENCE 65 AA; 6839 MW; 86269E85D08E4CB5 CRC64;

Query Match 100.0%; Score 18; DB 2; Length 65;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 ||||  
 DB 26 NLAA 29

## RESULT 13

O9EOK3 PRELIMINARY; PRT; 65 AA.  
 AC O9EOK3; (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE KRUPPEL-RELATED ZINC FINGER PROTEIN F80-L.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=ADIPOSE;  
 RA Bl S., Tartaglia L.A., Reisman M.L.;  
 RT "Isolation of a cDNA clone for a novel Kruppel-related zinc finger  
 protein from mouse adipose tissue."  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF211868; AAC43565.1; -  
 SQ SEQUENCE 65 AA; 7837 MW; 6BA6AD3649A4FA CRC64;

Query Match 100.0%; Score 18; DB 11; Length 65;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 ||||  
 DB 39 NLAA 42

## RESULT 14

O9OJ53 PRELIMINARY; PRT; 68 AA.  
 AC O9OJ53;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE STRAIN 229, COMPLETE GENOME.  
 GN U6.  
 OS Human herpesvirus 6B.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Roseolovirus.  
 OX NCBI\_TaxID=32604;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=229;  
 RA MEDLINE=99412318; PubMed=10482553;  
 RA Dominguez G., Dambaugh T.R., Stamey F.R., Dewhurst S., Inoue N.,  
 RA Pellett P.E.;  
 RT "Human herpesvirus 6B genome sequence: coding content and comparison  
 with human herpesvirus 6A.";  
 RL J. Virol. 73:8040-8052(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=229;  
 RA Pellett P.E., Dominguez G., Dambaugh T.R., Stamey F.R., Dewhurst S.,

RA Inoue N.;  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF157706; AAD49624.1; -  
 SQ SEQUENCE 68 AA; 7303 MW; E6050B80548D0491 CRC64;

Query Match 100.0%; Score 18; DB 14; Length 68;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 ||||  
 DB 3 NLAA 6

## RESULT 15

O9IOH0 PRELIMINARY; PRT; 69 AA.  
 AC O9IOH0;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE HYPOTHETICAL PROTEIN PA2668.  
 GN PA2668.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PA01;  
 RA MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Gardner R.L., Goultier S.N., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lapid G., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen."  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004695; AAG06056.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 69 AA; 7798 MW; 14B143130D08650 CRC64;

Query Match 100.0%; Score 18; DB 2; Length 69;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 ||||  
 DB 53 NLAA 56

Search completed: September 6, 2001, 16:49:57  
 Job time: 741 sec

```

OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reimach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvares R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Boye J.M., Brites M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Farto J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitzling J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peloto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Paes R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zaag M.A., Zatz M., Medanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL: AE003994; AAF84503.1; -.
KW Hypothetical protein.
SQ SEQUENCE 62 AA; 6116 MW; F6F27BD28CFC337 CRC64;

Query Match
Best Local Similarity 100.0%; Score 18; DB 2; Length 62;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLA 4
DB 22 NLA 25

RESULT 10
O98037 PRELIMINARY; PRT; 62 AA.
AC O98037;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE MAJOR HISTOCOMPATIBILITY COMPLEX CLASS II ANTIGEN (FRAGMENT).
GN MHC-MAHU-DOAL.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RA Sauermann U.;
RT "DQ-nucleotide analysis in Rhesus macaques: Implications for the
RT evolution of these genes.";
RL Tissue Antigens 0:0-0(1998).
DR EMBL: AF091334; AAC72829.1; -.
DR InterPro: IPR001003; -.
DR Pfam: PF00993; MHC_II_alpha; 1.
KW MHC.

```

```

FT NON_TER 1
FT NON_TER 62
SQ SEQUENCE 62 AA; 7070 MW; 5AB74BF5D82007EC CRC64;

Query Match
Best Local Similarity 100.0%; Score 18; DB 7; Length 62;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLA 4
DB 48 NLA 51

RESULT 11
O98141 PRELIMINARY; PRT; 62 AA.
AC O98141;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE MHC CLASS II ANTIGEN DO-ALPHA (FRAGMENT).
GN MHC-DOA.
OS Papio cynocephalus (Yellow baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Papio.
OX NCBI_TaxID=9556;
RN [1]
RP SEQUENCE FROM N.A.
RA Alberts S.C.;
RT "MHC-DOAL nucleotide sequences from two populations of baboons, Papio
RT cynocephalus cynocephalus and Papio cynocephalus papio.";
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF110834; AAC96333.1; -.
DR InterPro: IPR001003; -.
DR Pfam: PF00993; MHC_II_alpha; 1.
KW MHC.
FT NON_TER 1
FT NON_TER 62
SQ SEQUENCE 62 AA; 7188 MW; 40B2462AC824C23C CRC64;

Query Match
Best Local Similarity 100.0%; Score 18; DB 7; Length 62;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLA 4
DB 48 NLA 51

RESULT 12
O56027 PRELIMINARY; PRT; 65 AA.
AC O56027;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE SSPA (FRAGMENT).
GN SSPA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=SL1344;
RX MEDLINE=96342378; PubMed=8748032;
RA Hueck C.J., Hantman M.J., Bajaj V., Johnston C., Lee C.A.,
RA Miller S.I.;
RT "Salmonella typhimurium secreted invasion determinants are homologous
RT to Shigella ipa proteins.";

```

GN BH1743.  
OS Bacillus halodurans.  
OC Bacteria/Firmicutes: Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=86865;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-C-125 / JCM 9153;  
RA Takami H., Nakasone K., Takaki Y.;  
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL, AF001513; BAB05462.1;  
SQ SEQUENCE 44 AA; 5140 MW; 97C657A01A29EBDD CRC64;

Query Match 100.0%; Score 18; DB 2; Length 44;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
Db 15 NLAA 18

RESULT 7  
ID 09VJ78 PRELIMINARY; PRT; 56 AA.  
AC 09VJ78;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE CG15292. PROTEIN.  
GN CG15292.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY.  
RA MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanaides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang O., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abell J.F., Agbayani A., An H.-U., Andrews-Plannkoc C., Baldwin D.,  
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Butts K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Flocker C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jaitali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003642; AAF53377.1;  
DR FLYbase; FBgn0040977; CG15292.  
SQ SEQUENCE 56 AA; 5926 MW; ECFE0402469F031C CRC64;

Query Match 100.0%; Score 18; DB 5; Length 56;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
Db 9 NLAA 12

RESULT 8  
ID 035497 PRELIMINARY; PRT; 60 AA.  
AC 035497;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE NADH DEHYDROGENASE SUBUNIT 2 (FRAGMENT).  
GN ND2.  
OS Paracentrotus lividus (Common sea urchin).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidea; Euechinozoa; Echinacea; Echinoida; Echinidae;  
OC Paracentrotus.  
OX NCBI\_TaxID=7656;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA MEDLINE=87248108; PubMed=3596250;  
RA Cantaloro P., Roberti M., Morisco P., Rainaldi G., Gadaleta M.N.,  
RA Saccone C.;  
RT "A novel gene order in the Paracentrotus lividus mitochondrial  
RT genome."  
RL gene 53:41-54(1987).  
CC -1- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS  
CC CHAINS.  
CC EMBL; M16516; AAA31987.2; -  
DR InterPro; IPR001750; -  
DR Pfam; PF00361; oxidored\_q1; 1.  
KW Mitochondrion; NAD; Oxidoreductase.  
FT NON\_TER 1  
FT NON\_TER 60  
FT NON\_TER 60  
SQ SEQUENCE 60 AA; 6710 MW; ADDB99773497BFF4 CRC64;

Query Match 100.0%; Score 18; DB 8; Length 60;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
Db 22 NLAA 25

RESULT 9  
ID 09PCT4 PRELIMINARY; PRT; 62 AA.  
AC 09PCT4;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE HYPOTHEMETICAL PROTEIN XF1694.  
GN XF1694.  
OS Xylella fastidiosa.

OS Leptospira interrogans.  
 OC Bacteria; Spirochaetales; Leptospiaceae; Leptospira.  
 OX NCBI\_TaxID=173;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=92325069; PubMed=1624463;  
 RA Trieba G.A., Bolln C.A., Zuercher R.L.;  
 RT "Characterization of the periplasmic flagellum proteins of Leptospira  
 interrogans.";  
 RL J. Bacteriol. 174:4761-4768(1992).  
 SO SEQUENCE 30 AA; 3464 MW; 3BDEF3FDSCA4969 CRC64;

Query Match 100.0%; Score 18; DB 2; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 DB 6 NLAA 9

RESULT 3  
 O9S6T2 PRELIMINARY; PRT; 31 AA.  
 AC O9S6T2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE HYPOTHETICAL 3.3 KDA PROTEIN (FRAGMENT).  
 GN SCF134.21.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Saunders D., Harris D.;  
 RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kleser H.M., Denaplatte D., Elchner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL; AL049661; CAB41219.1; -.  
 KW Hypothetical protein.  
 FT NON-TER 31  
 SO SEQUENCE 31 AA; 3290 MW; 293703A3C0CE1A56 CRC64;

Query Match 100.0%; Score 18; DB 2; Length 31;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 DB 28 NLAA 31

RESULT 4  
 O9RIH7 PRELIMINARY; PRT; 37 AA.  
 AC O9RIH7;

DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE INSULIN RECEPTOR PRECURSOR (FRAGMENT).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY;  
 RA Liu Y., Tam J.W.O.;  
 RT "Partial sequence of rat insulin receptor gene.";  
 RL Submitted (Nov-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF110200; AAD40866.1; -.  
 DR EMBL; AF110199; AAD40866.1; JOINED.  
 KW Receptor.  
 FT NON-TER 1  
 FT NON-TER 37  
 SO SEQUENCE 37 AA; 3966 MW; 04FFA17F3B1B5855 CRC64;

Query Match 100.0%; Score 18; DB 11; Length 37;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 DB 13 NLAA 16

RESULT 5  
 O53518 PRELIMINARY; PRT; 42 AA.  
 AC O53518;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE UREG PROTEIN (FRAGMENT).  
 GN UREG.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92121127; PubMed=1310093;  
 RA Collins C.M., Gutman D.M.;  
 RT "Insertional inactivation of an Escherichia coli urease gene by  
 IS3411.";  
 RL J. Bacteriol. 174:883-888(1992).  
 DR EMBL; S77596; AAB21161.1; -.  
 FT NON-TER 1  
 SO SEQUENCE 42 AA; 4364 MW; FCC09C2DC270E9AC CRC64;

Query Match 100.0%; Score 18; DB 2; Length 42;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 DB 14 NLAA 17

RESULT 6  
 O9KC31 PRELIMINARY; PRT; 44 AA.  
 AC O9KC31;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE BH1743 PROTEIN.

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:56 ; Search time 231.42 Seconds  
(without alignments)  
2.287 Million cell updates/sec

Title: US-09-603-713-29

Perfect score: 18

Sequence: 1 NLA 4

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL.16:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp.unclassified:\*  
13: sp.vertebrate:\*  
14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	100.0	27	10	Q9S8R3	Q9S8R3 brassica nap
2	18	100.0	30	2	Q9R5K3	Q9R5K3 leptospira
3	18	100.0	31	2	Q9S6T2	Q9S6T2 streptomyc
4	18	100.0	37	11	Q9RIH7	Q9RIH7 rattus norv
5	18	100.0	42	2	Q53518	Q53518 escherichia
6	18	100.0	44	2	Q9K31	Q9K31 bacillus ba
7	18	100.0	56	5	Q9VJ78	Q9VJ78 drosophila
8	18	100.0	60	8	Q35497	Q35497 paracentrot
9	18	100.0	62	2	Q9PCT4	Q9PCT4 xyella fas
10	18	100.0	62	7	Q9R8J7	Q9R8J7 macaca mula
11	18	100.0	62	7	Q9R8J7	Q9R8J7 papio cynoc
12	18	100.0	65	2	Q56027	Q56027 salmone
13	18	100.0	65	11	Q9R5K3	Q9R5K3 mus musculi
14	18	100.0	68	14	Q9O553	Q9O553 human herpe
15	18	100.0	69	2	Q9I0H0	Q9I0H0 pseudomonas
16	18	100.0	73	4	Q9G2R6	Q9G2R6 homo sapien
17	18	100.0	75	2	Q50152	Q50152 mycobacteri
18	18	100.0	78	5	Q23756	Q23756 callinectes
19	18	100.0	81	5	Q9XWY9	Q9XWY9 caenorhabdi

20	18	100.0	82	14 Q69546	Q69546 human herpe
21	18	100.0	83	9 Q9ZXT5	Q9ZXT5 bacterioph
22	18	100.0	84	5 Q9G525	Q9G525 ancylostoma
23	18	100.0	84	14 Q9IKV2	Q9IKV2 human limun
24	18	100.0	84	14 Q9IKV2	Q9IKV2 human limun
25	18	100.0	84	14 Q9P236	Q9P236 xestia c-nl
26	18	100.0	85	14 Q9WT58	Q9WT58 human herpe
27	18	100.0	89	10 Q41446	Q41446 solanum tub
28	18	100.0	89	14 Q85559	Q85559 baboon endo
29	18	100.0	90	2 Q49060	Q49060 mycoplasma
30	18	100.0	91	1 Q30318	Q30318 archaeoglob
31	18	100.0	92	10 Q9S9G1	Q9S9G1 brassica na
32	18	100.0	93	3 Q14385	Q14385 schistosach
33	18	100.0	93	9 Q9ZX32	Q9ZX32 mycobacteri
34	18	100.0	96	10 P93148	P93148 glycyrrhiza
35	18	100.0	97	6 Q9N156	Q9N156 papio hamad
36	18	100.0	97	6 Q9N0P2	Q9N0P2 ceropithec
37	18	100.0	97	10 Q8S9G8	Q8S9G8 lycopersico
38	18	100.0	97	10 Q9LD79	Q9LD79 juniperus v
39	18	100.0	102	10 Q9M4Y4	Q9M4Y4 betula verr
40	18	100.0	103	1 Q33157	Q33157 methanosarc
41	18	100.0	103	6 Q9TV18	Q9TV18 bos taurus
42	18	100.0	105	2 Q53190	Q53190 rhodobacter
43	18	100.0	106	2 Q9RIA2	Q9RIA2 streptomyce
44	18	100.0	106	11 Q9EOK2	Q9EOK2 mus musculu
45	18	100.0	107	2 Q9KFX1	Q9KFX1 vibrio chol

## ALIGNMENTS

RESULT 1  
Q9S8R3 PRELIMINARY; PRT: 27 AA.  
ID Q9S8R3;  
AC Q9S8R3;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE ACYL-BINDING/LIPID-TRANSFER PROTEIN (FRAGMENT).  
OS Brassica napus (Rape).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Fossidae; eurosids II;  
OC Brassicales; Brassicaceae; Brassica.  
OX NCBI\_Taxid=3708;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=94002198; PubMed=8399333;  
FA Ostergaard J., Veronille C., Schoentgen F., Kader J.C.;  
RT "acyl-binding/lipid-transfer proteins from rape seedlings, a novel  
category of proteins interacting with lipids.";  
RL Biochim. Biophys. Acta 1170:109-117(1993).  
DR InterPro: IPR000528; LRP; 1.  
DR Pfam: PF00275; LRP; 1.  
SQ SEQUENCE 27 AA; 2634 MW; 705628C09E5BFE6B CRC64;

Query Match 100.0%; Score 18; DB 10; Length 27;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLA 4  
DB 10 NLA 13  
RESULT 2  
Q9R5K3 PRELIMINARY; PRT: 30 AA.  
ID Q9R5K3;  
AC Q9R5K3;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE 34 KDA PERIPLASMIC FLAGELLA CORE PROTEIN (FRAGMENT).

```

XX    US5968757-A.
PM
XX    19-OCT-1999.
PD
XX    23-FEB-1998;    98US-0027536.
PF
XX    26-JAN-1996;    96US-0592696.
PR
XX    (UYVA-) UNIV VANDERBILT.
PA
XX
XX    Powers AC;
PI
XX
XX    WPI: 1999-590400/50.
DR
XX
XX    Assay for antibodies to glutamic acid decarboxylase useful for
PT    diagnosis of diabetes or risk of diabetes
XX
XX    Disclosure; Column 23-26; 16pp; English.
PS
XX
XX    This invention describes a novel assay for antibodies to glutamic acid
CC    decarboxylase (GAD) in a sample. The method of the invention comprises
CC    (a) contacting the sample with a chimeric polypeptide comprising: (i) an
CC    epitope present in amino acids 270-435, 451-570, 221-585, 270-442,
CC    443-513, 528-585, 451-585 or 221-442 of GAD65 protein; and (ii) a
CC    structural region comprising a polypeptide from a GAD67 protein; and (b)
CC    detecting any binding between antibodies in the sample and the chimeric
CC    polypeptide. The assay is useful in screening methods to determine if
CC    subjects are at risk of developing insulin-dependent diabetes mellitus
CC    (IDDM) and if diabetic subjects have type I diabetes. The chimeric
CC    polypeptide is a more specific diagnostic for insulin-dependent diabetes
CC    mellitus (IDDM) than intact GAD65 and produces fewer false positives than
CC    intact GAD65. This sequence represents a chimeric rat GAD65/human GAD67
CC    fusion protein epitope described in the invention.
XX
SQ    Sequence    166 AA:

```

```

Query Match          71.7%; Score 33; DB 20; Length 166;
Best Local Similarity 66.7%; Pred. No. 46;
Matches    6; Conservative    2; Mismatches    1; Indels    0; Gaps    0;
QY    2 GSGVLSRK 10
      | | : | : | |
DB    99 GSGVLSRK 107

```

Search completed: September 6, 2001, 16:43:32  
 Job time: 361 sec





PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 73.9%; Score 34; DB 21; Length 477;  
 Best Local Similarity 87.5%; Pred. No. 92;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGVILS 8  
 DB 381 Vsgvils 388

RESULT 11  
 AAR76655  
 ID AAR76655 standard; peptide: 14 AA.  
 AC AAR76655;

XX 05-MAR-1996 (first entry)

DE Peptide derived from human glutamic acid decarboxylase 14.

XX diabetes; T-cell subpopulation; detection; antigen production;  
 KM diagnosis; autoimmune disease.

OS Homo sapiens.

XX DE4418091-A1.

XX 27-JUL-1995.

XX 24-MAY-1994; 94DE-4418091.

XX 04-FEB-1994; 94DE-4403522.

XX 20-JAN-1994; 94DE-4401629.

XX (ENDL/) ENDL J.  
 PA (BOEF) BOEHRINGER MANNHEIM GMBH.

XX Albert W, Dormair K, Endl J, Jung G, Meinel E;  
 PI Stahl P, Schendel D;

XX WPI; 1995-264505/35.

XX Antigen-specific activated T-lymphocytes and their detection - by  
 PT interaction with inventive peptide(s) of peptide-MHC complexes;  
 PT useful in diagnosis of e.g. diabetes and autoimmune diseases

XX Claim 1; Fig 2; 21pp; German.

XX AAR76642-62 are derived from human glutamic acid decarboxylase and  
 CC specifically react with T-cell sub-populations isolated from recently  
 CC diagnosed Type-I diabetes. Pharmaceutical compns. contg. these  
 CC peptides and those shown in AAR7571-72, are useful for the diagnosis of

CC a disease or predispositions of immune system diseases, tumours, and  
 CC autoimmune diseases, including diabetes. The peptides are able to detect  
 CC specific T-cell subpopulations that are then used for antigen prodn.,  
 CC e.g. by reinfection.

XX Sequence 14 AA;

Query Match 71.7%; Score 33; DB 16; Length 14;  
 Best Local Similarity 66.7%; Pred. No. 3;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSGVILSRK 10  
 DB 2 gsgvilmstr 10

RESULT 12  
 AAM18863  
 ID AAM18863 standard; peptide: 14 AA.  
 AC AAM18863;

XX 05-JAN-1998 (first entry)

DE 65 KD Glutamic acid decarboxylase peptide fragment 14.

XX GAD; 65 kd; human; glutamic acid decarboxylase; autoractive; diagnosis;  
 KM insulin-dependent diabetes mellitus; IDDM; cell-mediated disease;  
 KM predisposition; autoimmune; tumour; rheumatoid arthritis;  
 XX multiple sclerosis.

OS Synthetic.

XX DE19526561-A1.

XX 23-JAN-1997.

XX 20-JUL-1995; 95DE-1026561.

XX 20-JUL-1995; 95DE-1026561.

XX (BOEF) BOEHRINGER MANNHEIM GMBH.

XX Donie F, Endl J, Ganz M, Jung G, Kientsch-engel R;  
 PI Pozzilli P, Stahl P;

XX WPI; 1997-088254/09.

XX Skin test for diagnosis of cell-mediated diseases, esp. diabetes -  
 PT involving intradermal admin. of auto-reactive substances

XX Claim 11; Fig 2; 12pp; German.

XX AAM18842-70 are peptide fragments of the 65 kd human glutamic acid  
 CC decarboxylase (GAD). The fragments are autoreactive substances used for  
 CC diagnosis of insulin-dependent diabetes mellitus (IDDM). The diagnosis  
 CC is determined by using a claimed method for diagnosis of cell-mediated  
 CC diseases or a predisposition to cell-mediated diseases, which is  
 CC effected by administering an autoreactive substance intradermally and  
 CC establishing the diagnosis on the basis of the occurrence or lack of a  
 CC positive reaction at the site of administration. The method is used for  
 CC diagnosis of autoimmune and tumour diseases, preferably T-cell-mediated  
 CC diseases such as rheumatoid arthritis, multiple sclerosis and especially  
 CC IDDM.

XX Sequence 14 AA;

Query Match 71.7%; Score 33; DB 18; Length 14;  
 Best Local Similarity 66.7%; Pred. No. 3;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0138763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 29-JUN-1999; 99US-0141287.  
PR 30-JUN-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 01-JUL-1999; 99US-0142055.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142877.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 23-JUL-1999; 99US-0145276.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147260.  
PR 05-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 27-AUG-1999; 99US-0151303.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151308.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154479.  
PR 20-SEP-1999; 99US-0155139.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155486.  
PR 28-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 04-OCT-1999; 99US-0157753.  
PR 05-OCT-1999; 99US-0157865.  
PR 06-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158232.  
PR 08-OCT-1999; 99US-0158369.  
PR 12-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159320.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.

PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 73.9%; Score 34; DB 21; Length 471;  
Best Local Similarity 87.5%; Pred. No. 91;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VCSGVLLS 8  
Db 375 vsgsvlls 382

RESULT 10  
AAG36902  
ID AAG36902 standard; Protein; 477 AA.  
XX  
AC AAG36902;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45289.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hydridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.

PR 29-SEP-1999: 99US-0156596.  
 PR 04-OCT-1999: 99US-0157111.  
 PR 05-OCT-1999: 99US-0157753.  
 PR 06-OCT-1999: 99US-0157865.  
 PR 07-OCT-1999: 99US-0158029.  
 PR 08-OCT-1999: 99US-0158233.  
 PR 12-OCT-1999: 99US-0158366.  
 PR 13-OCT-1999: 99US-0159293.  
 PR 13-OCT-1999: 99US-0159294.  
 PR 13-OCT-1999: 99US-0159295.  
 PR 14-OCT-1999: 99US-0159329.  
 PR 14-OCT-1999: 99US-0159330.  
 PR 14-OCT-1999: 99US-0159331.  
 PR 14-OCT-1999: 99US-0159637.  
 PR 14-OCT-1999: 99US-0159638.  
 PR 18-OCT-1999: 99US-0159584.  
 PR 21-OCT-1999: 99US-0160741.  
 PR 21-OCT-1999: 99US-0160767.  
 PR 21-OCT-1999: 99US-0160768.  
 PR 21-OCT-1999: 99US-0160770.  
 PR 21-OCT-1999: 99US-0160814.  
 PR 21-OCT-1999: 99US-0160815.  
 PR 22-OCT-1999: 99US-0160980.  
 PR 22-OCT-1999: 99US-0160981.  
 PR 22-OCT-1999: 99US-0160988.  
 PR 25-OCT-1999: 99US-0161404.  
 PR 25-OCT-1999: 99US-0161405.  
 PR 25-OCT-1999: 99US-0161406.  
 PR 25-OCT-1999: 99US-0161350.  
 PR 26-OCT-1999: 99US-0161359.  
 PR 26-OCT-1999: 99US-0161360.  
 PR 26-OCT-1999: 99US-0161361.  
 PR 28-OCT-1999: 99US-0161920.  
 PR 28-OCT-1999: 99US-0161932.  
 PR 28-OCT-1999: 99US-0161993.  
 PR 29-OCT-1999: 99US-0162142.

Query Match 73.9% Score 34: DB 21: Length 442;  
 Best Local Similarity 87.5% Pred. No. 85;  
 Matches 7: Conservative 1: Mismatches 0: Indels 0: Gaps 0;

OY 1 VSGGLVLS 8  
 DB 346 vsggvlvs 353

RESULT 9  
 AAC36903  
 ID AAC36903 standard; Protein: 471 AA.  
 AC AAC36903;  
 XX  
 DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 45290.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 hybridisation assay; genetic mapping; gene expression control; promoter;  
 termination sequence.

OS Arabidopsis thaliana.

XX  
 PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999: 99US-0121825.  
 PR 05-MAR-1999: 99US-0123180.

PR 09-MAR-1999: 99US-0123548.  
 PR 23-MAR-1999: 99US-0125788.

PR 25-MAR-1999: 99US-0126266.  
 PR 29-MAR-1999: 99US-0126785.

PR 01-APR-1999: 99US-0127462.  
 PR 06-APR-1999: 99US-0128234.  
 PR 08-APR-1999: 99US-0128714.  
 PR 16-APR-1999: 99US-0129845.  
 PR 19-APR-1999: 99US-0130077.  
 PR 21-APR-1999: 99US-0130449.  
 PR 23-APR-1999: 99US-0130510.  
 PR 23-APR-1999: 99US-0130891.  
 PR 28-APR-1999: 99US-0131449.  
 PR 30-APR-1999: 99US-0132048.  
 PR 30-APR-1999: 99US-0132407.  
 PR 04-MAY-1999: 99US-0132484.  
 PR 05-MAY-1999: 99US-0132485.  
 PR 06-MAY-1999: 99US-0132486.  
 PR 06-MAY-1999: 99US-0132487.  
 PR 07-MAY-1999: 99US-0132863.  
 PR 11-MAY-1999: 99US-0134256.  
 PR 14-MAY-1999: 99US-0134218.  
 PR 14-MAY-1999: 99US-0134219.  
 PR 14-MAY-1999: 99US-0134221.  
 PR 14-MAY-1999: 99US-0134370.  
 PR 18-MAY-1999: 99US-0134768.  
 PR 19-MAY-1999: 99US-0134941.  
 PR 20-MAY-1999: 99US-0135124.  
 PR 21-MAY-1999: 99US-0135353.  
 PR 24-MAY-1999: 99US-0135629.  
 PR 25-MAY-1999: 99US-0136021.  
 PR 27-MAY-1999: 99US-0136392.  
 PR 28-MAY-1999: 99US-0136782.  
 PR 01-JUN-1999: 99US-0137222.  
 PR 03-JUN-1999: 99US-0137528.  
 PR 04-JUN-1999: 99US-0137502.  
 PR 07-JUN-1999: 99US-0137724.  
 PR 08-JUN-1999: 99US-0138094.  
 PR 10-JUN-1999: 99US-0138540.  
 PR 10-JUN-1999: 99US-0138847.  
 PR 14-JUN-1999: 99US-0139119.  
 PR 16-JUN-1999: 99US-0139452.  
 PR 16-JUN-1999: 99US-0139453.  
 PR 17-JUN-1999: 99US-0139492.  
 PR 18-JUN-1999: 99US-0139454.  
 PR 18-JUN-1999: 99US-0139455.  
 PR 18-JUN-1999: 99US-0139456.  
 PR 18-JUN-1999: 99US-0139457.  
 PR 18-JUN-1999: 99US-0139458.  
 PR 18-JUN-1999: 99US-0139459.  
 PR 18-JUN-1999: 99US-0139460.  
 PR 18-JUN-1999: 99US-0139461.  
 PR 18-JUN-1999: 99US-0139462.  
 PR 18-JUN-1999: 99US-0139463.  
 PR 18-JUN-1999: 99US-0139750.  
 PR 18-JUN-1999: 99US-0139763.  
 PR 21-JUN-1999: 99US-0139817.  
 PR 22-JUN-1999: 99US-0139899.  
 PR 23-JUN-1999: 99US-0140353.  
 PR 23-JUN-1999: 99US-0140354.  
 PR 24-JUN-1999: 99US-0140695.  
 PR 28-JUN-1999: 99US-0140823.  
 PR 29-JUN-1999: 99US-0140991.  
 PR 30-JUN-1999: 99US-0141287.  
 PR 01-JUL-1999: 99US-0141842.  
 PR 01-JUL-1999: 99US-0142154.  
 PR 02-JUL-1999: 99US-0142055.  
 PR 06-JUL-1999: 99US-0142390.  
 PR 08-JUL-1999: 99US-0142803.  
 PR 09-JUL-1999: 99US-0142920.  
 PR 12-JUL-1999: 99US-0142977.  
 PR 13-JUL-1999: 99US-0143542.  
 PR 14-JUL-1999: 99US-0143624.  
 PR 15-JUL-1999: 99US-0144005.  
 PR 16-JUL-1999: 99US-0144085.  
 PR 16-JUL-1999: 99US-0144086.  
 PR 19-JUL-1999: 99US-0144325.

XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0128845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131448.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140923.  
 PR 29-JUN-1999; 99US-0140921.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 05-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147933.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149725.  
 PR 23-AUG-1999; 99US-0149925.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0152370.  
 PR 13-SEP-1999; 99US-0153738.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.

PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158025.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	23-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161982.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

AC	AGG41662;		PR	18-JUN-1999;	990S-0139460.
XX			PR	18-JUN-1999;	990S-0139750.
DT	18-OCT-2000 (first entry)		PR	18-JUN-1999;	990S-0139763.
XX			PR	21-JUN-1999;	990S-0139817.
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 51865.		PR	22-JUN-1999;	990S-0139899.
DE			PR	23-JUN-1999;	990S-0140353.
XX			PR	23-JUN-1999;	990S-0140354.
XX	Protein identification: signal transduction pathway; metabolic pathway;		PR	24-JUN-1999;	990S-0140695.
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		PR	24-JUN-1999;	990S-0140695.
KM	termination sequence.		PR	28-JUN-1999;	990S-0140823.
XX			PR	29-JUN-1999;	990S-0140991.
XX	Arabidopsis thaliana.		PR	30-JUN-1999;	990S-0141287.
OS			PR	01-JUL-1999;	990S-0141842.
PN	EP1033405-A2.		PR	01-JUL-1999;	990S-0142154.
XX			PR	02-JUL-1999;	990S-0142055.
XX			PR	06-JUL-1999;	990S-0142390.
PD	06-SEP-2000.		PR	08-JUL-1999;	990S-0142803.
XX			PR	09-JUL-1999;	990S-0142920.
PF	25-FEB-2000; 2000EP-0301439.		PR	12-JUL-1999;	990S-0142977.
XX			PR	13-JUL-1999;	990S-0143542.
XX			PR	13-JUL-1999;	990S-0143542.
PR	25-FEB-1999;	990S-0121825.	PR	14-JUL-1999;	990S-0143624.
PR	05-MAR-1999;	990S-0123180.	PR	15-JUL-1999;	990S-0144005.
PR	09-MAR-1999;	990S-0123549.	PR	16-JUL-1999;	990S-0144085.
PR	23-MAR-1999;	990S-0125788.	PR	16-JUL-1999;	990S-0144334.
PR	25-MAR-1999;	990S-0126264.	PR	19-JUL-1999;	990S-0144335.
PR	29-MAR-1999;	990S-0126785.	PR	16-JUL-1999;	990S-0144326.
PR	01-APR-1999;	990S-0127462.	PR	16-JUL-1999;	990S-0144326.
PR	06-APR-1999;	990S-0128234.	PR	19-JUL-1999;	990S-0144331.
PR	08-APR-1999;	990S-0128714.	PR	19-JUL-1999;	990S-0144332.
PR	16-APR-1999;	990S-0129845.	PR	19-JUL-1999;	990S-0144334.
PR	19-APR-1999;	990S-0130077.	PR	19-JUL-1999;	990S-0144335.
PR	21-APR-1999;	990S-0130449.	PR	20-JUL-1999;	990S-0144332.
PR	23-APR-1999;	990S-0130510.	PR	20-JUL-1999;	990S-0144632.
PR	23-APR-1999;	990S-0130891.	PR	20-JUL-1999;	990S-0144884.
PR	28-APR-1999;	990S-0131049.	PR	21-JUL-1999;	990S-0144881.
PR	30-APR-1999;	990S-0132048.	PR	21-JUL-1999;	990S-0145086.
PR	30-APR-1999;	990S-0132407.	PR	21-JUL-1999;	990S-0145088.
PR	04-MAY-1999;	990S-0132484.	PR	22-JUL-1999;	990S-0145085.
PR	05-MAY-1999;	990S-0132485.	PR	22-JUL-1999;	990S-0145087.
PR	06-MAY-1999;	990S-0132486.	PR	22-JUL-1999;	990S-0145089.
PR	07-MAY-1999;	990S-0132487.	PR	22-JUL-1999;	990S-0145129.
PR	11-MAY-1999;	990S-0132863.	PR	23-JUL-1999;	990S-0145118.
PR	14-MAY-1999;	990S-0134218.	PR	23-JUL-1999;	990S-0145218.
PR	14-MAY-1999;	990S-0134219.	PR	23-JUL-1999;	990S-0145224.
PR	14-MAY-1999;	990S-0134221.	PR	26-JUL-1999;	990S-0145276.
PR	14-MAY-1999;	990S-0134370.	PR	27-JUL-1999;	990S-0145918.
PR	18-MAY-1999;	990S-0134768.	PR	27-JUL-1999;	990S-0145919.
PR	19-MAY-1999;	990S-0134941.	PR	28-JUL-1999;	990S-0145951.
PR	20-MAY-1999;	990S-0135124.	PR	02-AUG-1999;	990S-0146388.
PR	21-MAY-1999;	990S-0135353.	PR	02-AUG-1999;	990S-0146388.
PR	24-MAY-1999;	990S-0135621.	PR	02-AUG-1999;	990S-0146389.
PR	25-MAY-1999;	990S-0136021.	PR	03-AUG-1999;	990S-0147038.
PR	27-MAY-1999;	990S-0136392.	PR	04-AUG-1999;	990S-0147204.
PR	28-MAY-1999;	990S-0136782.	PR	04-AUG-1999;	990S-0147302.
PR	01-JUN-1999;	990S-0137222.	PR	05-AUG-1999;	990S-0147192.
PR	03-JUN-1999;	990S-0137528.	PR	05-AUG-1999;	990S-0147260.

PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139899.  
PR 22-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141827.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143542.  
PR 13-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0143604.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 21-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145145.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145813.  
PR 27-JUL-1999; 99US-0145818.  
PR 27-JUL-1999; 99US-0145919.  
PR 27-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151438.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159283.  
PR 13-OCT-1999; 99US-0159284.  
PR 13-OCT-1999; 99US-0159285.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161892.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 76.1%; Score 35; DB 21; Length 1079;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSCVLLSRK 10  
1:1:1:1:1:1  
DB 790 gngllvsrk 798

RESULT 6  
AACG1662  
ID AACG1662 standard; Protein: 1132 AA.  
XX



PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157863.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 76.1%; Score 35; DB 21; Length 1072;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 GSGVLSRK 10  
|:|:|:|:|

DB 783 gngllsvrk 791  
RESULT 5  
AAG41663  
ID AAG41663 standard; Protein: 1079 AA.  
XX AAG41663;  
AC  
XX 18-OCT-2000 (first entry)  
DT  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51866.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
XX  
PD 06-SEP-2000.  
XX  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0123788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.

Query Match	89.1%	Score 41	DB 13	Length 1078
Best Local Similarity	90.0%	Pred. No. 9/2		
Matches 9	Conservative	0	Mismatches	1
			Indels	0
			Gaps	0
QY	1 VGGYLLSRK 10			
DB	297 VGGYLLSRK 306			
RESULT 4				
ID	AA041664			
AC	AA041664 standard; Protein: 1072 AA.			
DT	18-OCT-2000 (first entry)			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 51867.			
KW	Protein identification; signal transduction pathway; metabolic pathway;			
KW	hybridisation assay; genetic mapping; gene expression control; promoter;			
KW	termination sequence.			
OS	Arabidopsis thaliana.			
XX	EPI033405-A2.			
XX	06-SEP-2000.			
PE	25-FEB-2000; 2000EP-0301439.			
XX	25-FEB-1999; 9905-0121825.			
XX	05-MAR-1999; 9905-0123180.			
PR	09-MAR-1999; 9905-0123548.			
PR	23-MAR-1999; 9905-0125788.			
PR	25-MAR-1999; 9905-0126264.			
PR	29-MAR-1999; 9905-0126785.			
PR	01-APR-1999; 9905-0127462.			
PR	06-APR-1999; 9905-0128334.			
PR	08-APR-1999; 9905-0128714.			
PR	16-APR-1999; 9905-0129845.			
PR	19-APR-1999; 9905-0130077.			
PR	21-APR-1999; 9905-0130449.			
PR	23-APR-1999; 9905-0130510.			
PR	28-APR-1999; 9905-0130891.			
PR	30-APR-1999; 9905-0131449.			
PR	04-MAY-1999; 9905-0132407.			
PR	05-MAY-1999; 9905-0132484.			
PR	06-MAY-1999; 9905-0132485.			
PR	06-MAY-1999; 9905-0132486.			
PR	07-MAY-1999; 9905-0132487.			
PR	11-MAY-1999; 9905-0134218.			
PR	14-MAY-1999; 9905-0134219.			
PR	14-MAY-1999; 9905-0134221.			
PR	14-MAY-1999; 9905-0134370.			
PR	18-MAY-1999; 9905-0134768.			
PR	19-MAY-1999; 9905-0134941.			
PR	20-MAY-1999; 9905-0135124.			
PR	21-MAY-1999; 9905-0135553.			
PR	24-MAY-1999; 9905-0135629.			
PR	25-MAY-1999; 9905-0136021.			
PR	27-MAY-1999; 9905-0136092.			
PR	28-MAY-1999; 9905-0136182.			
PR	01-JUN-1999; 9905-0137222.			
PR	03-JUN-1999; 9905-0137528.			
PR	04-JUN-1999; 9905-0137722.			
PR	07-JUN-1999; 9905-0137724.			
PR	08-JUN-1999; 9905-0138094.			
PR	10-JUN-1999; 9905-0138540.			

[illegible]

PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage  
 XX  
 XX Example 4; Page 33; 86pp; English.  
 CC The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 CC  
 XX Sequence 10 AA:

Query Match 100.0%; Score 46; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.0056;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLSRK 10  
 |||||  
 DB 1 vsgsvlstrk 10

## RESULT 2

AAB61350  
 ID AAB61350 standard; peptide; 10 AA.

AC AAB61350;

DT 02-APR-2001 (first entry)

XX Synthetic peptide #2.

DE Memapsin 2; catalyst; Alzheimer's.

XX Synthetic.

PN WO200100663-A2.

PD 04-JAN-2001.

PF 27-JUN-2000; 2000WO-US17661.

PR 28-JUN-1999; 99US-0141363.

PR 30-NOV-1999; 99US-0168060.

PR 25-JAN-2000; 2000US-017836.

PR 27-JAN-2000; 2000US-0178368.

PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

PI Tang JYN, Lin X, Koelsch G;

DR WPI: 2001-102885/11.

XX Purified recombinant catalytically active memapsin 2, used to screen

XX inhibitors of it, which are used to treat and prevent Alzheimer's

XX disease -

PS Disclosure: Page 86; 86pp; English.

XX The present invention relates to a purified recombinant

XX catalytically active memapsin 2. The invention may be used for

XX isolating inhibitors which are used to treat or prevent

XX Alzheimer's disease. The invention may also be used to screen

XX for individuals more genetically prone to develop Alzheimer's

XX disease.

XX  
 SO Sequence 10 AA:

Query Match 100.0%; Score 46; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.0056;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLSRK 10  
 |||||  
 DB 1 vsgsvlstrk 10

## RESULT 3

AAR28963  
 ID AAR28963 standard; Protein; 1078 AA.

AC AAR28963;

DT 01-APR-1993 (first entry)

XX Notch hn3k full length clone.

DE Human; Notch; plasmid; cDNA; clone; hn2k; expression library;

KW adhesion domain; Drosophila; epidermal growth factor; EGF;

KW EGF-like; homologous; repeat; EIR; EIR-11; EIR-12; PCR;

KW polymerase chain reaction; primer; cloning vector; Delta; Serrate;

KW neurogenic; topolytic; homotypic; heterotypic; differentiation;

KW quantitation; antibody.

XX Homo sapiens.

OS WO9219734-A.

PN 12-NOV-1992.

PD 01-MAY-1992; 92WO-US03651.

PF 03-MAY-1991; 91US-0695189.

PR 14-NOV-1991; 91US-0791923.

XX (INDV ) UNIV INDIANA FOUND.

PA (UYVA ) UNIV YALE.

XX Artavanis-Tsakonas S, Blaumueller CM, Fehon RG, Muskavitch MAT;

PI Rebay I, Shepard SB;

DR WPI: 1992-398861/48.

DR N-PSDB; AAQ30999.

XX Human Notch and Delta DNA and protein sequences - used for study

XX and manipulation of differentiation processes

XX Clalm 132; Fig 23; 239pp; English.

XX The sequence given represent the full length peptide sequence of

XX human Notch encoded by plasmid cDNA clone hn3k. A human expression

XX library was constructed and screening assays were carried out on the

XX basis of positive binding to the adhesion domain of Notch, ie. the

XX region that is homologous to Drosophila epidermal growth factor

XX (EGF)-like (homologous) repeat (ELR)-11 and -12. Alternatively the

XX sequence could be isolated by amplification using polymerase chain

XX reaction (PCR) primers. The isolated gene may be inserted into a

XX cloning vector and expressed. The Notch gene and also the Delta and

XX Serrate neurogenic genes are designated "topolytic" genes. The

XX proteins they encode are involved in specific homo- or heterotypic

XX interactions crucial to differentiation. The quantitation of mRNA

XX for human Notch and Delta and adhesive molecules, and study of its

XX expression are possible using the DNA and antibodies raised against

XX the Notch and Delta proteins.

XX Sequence 1078 AA:

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:32 : Search time 225.25 Seconds

(without alignments)  
2.691 Million cell updates/sec

Title: US-09-603-713-30

Perfect score: 46

Sequence: 1 VSGVLSRK 10

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: A.Geneseq.0601.\*  
2: /SIDS8/gcgdata/geneseq/AA1980.DAT.\*  
3: /SIDS8/gcgdata/geneseq/AA1981.DAT.\*  
4: /SIDS8/gcgdata/geneseq/AA1982.DAT.\*  
5: /SIDS8/gcgdata/geneseq/AA1983.DAT.\*  
6: /SIDS8/gcgdata/geneseq/AA1984.DAT.\*  
7: /SIDS8/gcgdata/geneseq/AA1985.DAT.\*  
8: /SIDS8/gcgdata/geneseq/AA1986.DAT.\*  
9: /SIDS8/gcgdata/geneseq/AA1987.DAT.\*  
10: /SIDS8/gcgdata/geneseq/AA1988.DAT.\*  
11: /SIDS8/gcgdata/geneseq/AA1989.DAT.\*  
12: /SIDS8/gcgdata/geneseq/AA1990.DAT.\*  
13: /SIDS8/gcgdata/geneseq/AA1991.DAT.\*  
14: /SIDS8/gcgdata/geneseq/AA1992.DAT.\*  
15: /SIDS8/gcgdata/geneseq/AA1993.DAT.\*  
16: /SIDS8/gcgdata/geneseq/AA1994.DAT.\*  
17: /SIDS8/gcgdata/geneseq/AA1995.DAT.\*  
18: /SIDS8/gcgdata/geneseq/AA1996.DAT.\*  
19: /SIDS8/gcgdata/geneseq/AA1997.DAT.\*  
20: /SIDS8/gcgdata/geneseq/AA1998.DAT.\*  
21: /SIDS8/gcgdata/geneseq/AA1999.DAT.\*  
22: /SIDS8/gcgdata/geneseq/AA2000.DAT.\*  
23: /SIDS8/gcgdata/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	100.0	10	22	AAH66588
2	46	100.0	10	22	AAH66588
3	41	89.1	1078	13	AAH61350
4	35	76.1	1072	21	AAH641664
5	35	76.1	1079	21	AAH641663
6	35	76.1	1132	21	AAH641662
7	34	73.9	11	21	AAH07256
8	34	73.9	442	21	AAH36504
9	34	73.9	471	21	AAH36503
10	34	73.9	477	21	AAH36502
11	33	71.7	14	16	AAH76055

12	33	71.7	14	18	AAH18863
13	33	71.7	37	20	AAH12603
14	33	71.7	166	20	AAH74718
15	33	71.7	166	20	AAH33657
16	33	71.7	166	21	AAH03073
17	33	71.7	169	21	AAH42394
18	33	71.7	181	13	AAH27218
19	33	71.7	181	13	AAH27219
20	33	71.7	341	15	AAH59522
21	33	71.7	341	15	AAH59523
22	33	71.7	371	20	AAH01300
23	33	71.7	371	20	AAH67443
24	33	71.7	431	21	AAH66203
25	33	71.7	463	21	AAH20250
26	33	71.7	502	13	AAH23645
27	33	71.7	540	13	AAH39520
28	33	71.7	540	13	AAH39521
29	33	71.7	544	15	AAH39524
30	33	71.7	544	15	AAH39525
31	33	71.7	554	15	AAH59518
32	33	71.7	554	15	AAH59519
33	33	71.7	584	19	AAH34519
34	33	71.7	584	19	AAH34520
35	33	71.7	584	20	AAH6017
36	33	71.7	584	20	AAH6018
37	33	71.7	585	13	AAH29628
38	33	71.7	585	13	AAH29629
39	33	71.7	585	13	AAH23644
40	33	71.7	585	13	AAH28756
41	33	71.7	585	15	AAH59516
42	33	71.7	585	15	AAH59517
43	33	71.7	585	16	AAH71641
44	33	71.7	585	16	AAH79105
45	33	71.7	585	18	AAH35361

## ALIGNMENTS

RESULT	1
AAH66588	standard; Peptide: 10 AA.
AAH66588;	
12-APR-2001	(first entry)
Synthetic memapsin 2 substrate peptide.	
memapsin 2; nootropic; neuroprotective; amyloid precursor protein;	
APP; memapsin 2 inhibitor; Alzheimer's disease.	
Synthetic.	
MO200100665-A2.	
04-JAN-2001.	
27-JUN-2000; 2000MO-US17742.	
28-JUN-1999; 99US-0141363.	
30-NOV-1999; 99US-0168060.	
25-JAN-2000; 2000US-017836.	
27-JAN-2000; 2000US-0178368.	
08-JUN-2000; 2000US-0210292.	
(OKLA-) OKLAHOMA MEDICAL RES FOUND.	
(UNIV ILLINOIS) FOUND.	
Tang JIN, Hong L, Ghosh AK.	
WPI; 2001-137933/14.	

65 kD glutamic aci  
Human 5' EST seque  
Amino acid sequenc  
Chimeric rat GAD65  
Rat GAD65 glutamic  
Human ORF2158  
Brain GAD. Homo s  
Islet GAD. Homo s  
GAD65 1-244 N-term  
GAD65 1-244 N-term  
Labile toxin (LT-B  
C. jejuni flagellin  
Nuclear transport  
Arabidopsis thaliana  
Human GAD65 gene p  
GAD65 1-45 N-term  
GAD65 1-45 N-term  
GAD65 345-385 C-te  
GAD65 345-385 C-te  
GAD65 1-31 deleted  
GAD65 1-31 deleted  
Human GAD65 protei  
Rat GAD65 protein.  
Human GAD65 protei  
Rat GAD 65. Ratu  
Human GAD 65. Hom  
Rat GAD65 gene pro  
Human pancreatic 1  
Human GAD65. Homo  
Rat GAD65. Ratus  
Human GAD. Homo s  
Human glutamic aci  
Human 65k-glutamic

APPLICANT: Musil, Roy  
TITLE OF INVENTION: Nucleic Acid Sequences And Expression  
TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From  
TITLE OF INVENTION: Flavobacterium heparinum  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hale and Dorr  
STREET: 1455 Pennsylvania Avenue, N.W.  
CITY: Washington, D.C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/258,639A  
FILING DATE: 10 JUNE 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Healey, William J.  
REGISTRATION NUMBER: 36,160  
REFERENCE/DOCKET NUMBER: 104385.116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)942-8400  
TELEFAX: (202)942-8484  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-258-639A-5

Query Match 100.0% Score 18; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4  
1111  
DB 7 NLAA 10

RESULT 12  
US-08-900-951-5  
Sequence 5, Application US/08900951  
Patent No. 5919693  
GENERAL INFORMATION:  
APPLICANT: Su, Hongsheng  
APPLICANT: Blain, Francoise  
APPLICANT: Bennett, Clark  
APPLICANT: Gu, Kangfu  
APPLICANT: Zimmermann, Joseph  
APPLICANT: Musil, Roy  
TITLE OF INVENTION: Nucleic Acid Sequences And Expression  
TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From  
TITLE OF INVENTION: Flavobacterium heparinum  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hale and Dorr  
STREET: 1455 Pennsylvania Avenue, N.W.  
CITY: Washington, D.C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/900,951

FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/258,639  
FILING DATE: 10 JUNE 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Healey, William J.  
REGISTRATION NUMBER: 36,160  
REFERENCE/DOCKET NUMBER: 104385.116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)942-8400  
TELEFAX: (202)942-8484  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-900-951-5

Query Match 100.0% Score 18; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4  
1111  
DB 7 NLAA 10

RESULT 13  
PCT-US95-07391A-5  
Sequence 5, Application PC/TUS9507391A  
GENERAL INFORMATION:  
APPLICANT: IBEX TECHNOLOGIES and  
APPLICANT: ZIMMERMANN, Joseph  
TITLE OF INVENTION: Nucleic Acid Sequences And Expression  
TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From  
TITLE OF INVENTION: Flavobacterium heparinum  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hale and Dorr  
STREET: 1455 Pennsylvania Avenue, N.W.  
CITY: Washington, D.C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07391A  
FILING DATE: 09-JUNE-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/258,639  
FILING DATE: 10 JUNE 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BAKER, Hollie L.  
REGISTRATION NUMBER: 31,321  
REFERENCE/DOCKET NUMBER: 104385.116PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)942-8400  
TELEFAX: (202)942-8484  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-07391A-5

Query Match 100.0%; Score 18; DB 5; Length 12;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
1111  
Db 7 NLAA 10

RESULT 14  
US-07-961-837-4  
Sequence 4, Application US/07961837  
Patent No. 5519116  
GENERAL INFORMATION:  
APPLICANT: WAGNER, THOMAS  
APPLICANT: OPPL, CRISTINA  
TITLE OF INVENTION: PEPTIDES REPRESENTING PERMUTATIONS OF  
NUMBER OF SEQUENCES: 9  
TITLE OF INVENTION: THE MASTOPARAN SEQUENCE  
CORRESPONDENCE ADDRESS:  
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,  
ADDRESS: P. C. Jefferson Davis Highway, Fourth Floor  
STREET: 1755 Jefferson Davis Highway, Fourth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/961,837  
FILING DATE: 19921016  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IT MI91 A 002770  
FILING DATE: 21-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5519116man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 3979-185-0X  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)413-3000  
TELEFAX: (703)413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-07-961-837-4

Query Match 100.0%; Score 18; DB 1; Length 14;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
1111  
Db 2 NLAA 5

RESULT 15  
US-08-232-453A-17  
Sequence 17, Application US/08232453A  
Patent No. 5589568  
GENERAL INFORMATION:  
APPLICANT: HIGASHIJIMA, TSUTOMU  
APPLICANT: ROSS, ELLIOTT M.

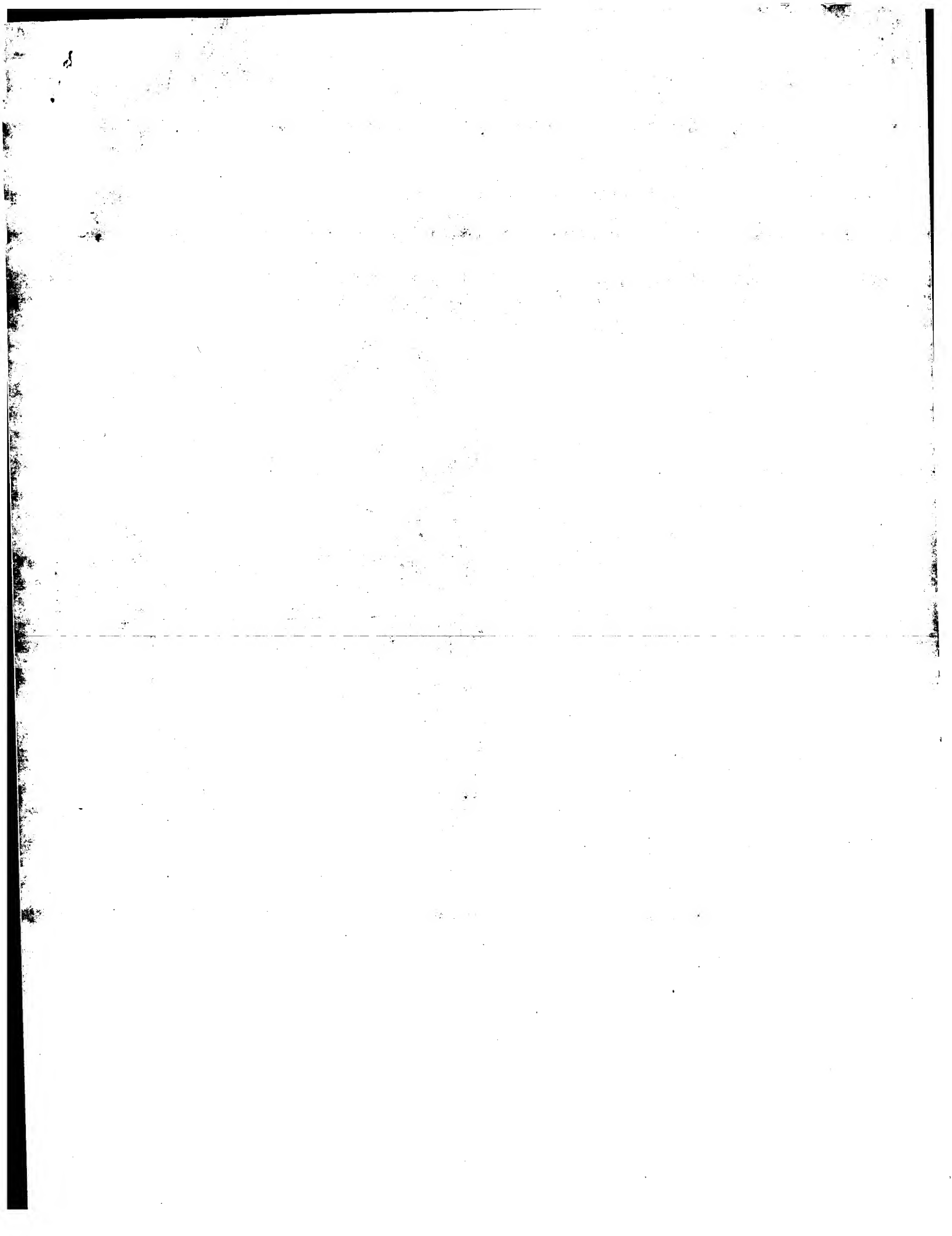
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
NUMBER OF INVENTION: MODULATING G PROTEIN ACTION  
NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRESS:  
ADDRESS: ARNOLD, WHITE & DUREE  
STREET: P.O. BOX 4433  
CITY: HOUSTON  
STATE: TX  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,453A  
FILING DATE: APRIL 22, 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/748,319  
FILING DATE: AUGUST 21, 1991  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: PARKER, DAVID L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: UTSD:253/PAR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-232-453A-17

Query Match 100.0%; Score 18; DB 1; Length 14;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
1111  
Db 2 NLAA 5

Search completed: September 6, 2001, 16:39:40  
Job time: 129 sec

---







PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage  
 XX  
 PS Disclosure: Page 84; 86pp; English.  
 CC  
 CC The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 CC  
 SQ Sequence 4 AA;  
 Query Match 100.0%; Score 18; DB 22; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 NLAA 4  
 Db 1 nlaa 4  
 RESULT 2  
 AAB61349  
 ID AAB61349 standard; peptide; 4 AA.  
 AC  
 AC AAB61349;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Synthetic peptide #1.  
 XX  
 XX Memapsin 2; catalyst; Alzheimer's.  
 KW  
 OS Synthetic.  
 OS  
 PN WO200100663-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 27-JUN-2000; 2000WO-US17661.  
 XX  
 PR 28-JUN-1999; 9905-0141363.  
 PR 30-NOV-1999; 9905-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX  
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PI  
 PI Tang JUN, Lin X, Koelsch G;  
 XX  
 DR WPI: 2001-102885/11.  
 XX  
 PT Purified recombinant catalytically active memapsin 2, used to screen  
 PT inhibitors of it, which are used to treat and prevent Alzheimer's  
 PT disease -  
 XX  
 PS Disclosure: Page 85; 86pp; English.  
 CC  
 CC The present invention relates to a purified recombinant  
 CC catalytically active memapsin 2. The invention may be used for  
 CC isolating inhibitors which are used to treat or prevent  
 CC Alzheimer's disease. The invention may also be used to screen  
 CC for individuals more genetically prone to develop Alzheimer's  
 CC disease.

XX  
 SQ Sequence 4 AA;  
 Query Match 100.0%; Score 18; DB 22; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 NLAA 4  
 Db 1 nlaa 4  
 RESULT 3  
 AAR55250  
 ID AAR55250 standard; Protein; 7 AA.  
 AC  
 AC AAR55250;  
 XX  
 DT 23-JAN-1995 (first entry)  
 XX  
 DE Highly preserved NMDA type receptor fragment for primer design.  
 XX  
 KW N-methyl-D-aspartic acid; glutamate receptors; NMDA;  
 KW nervous information transmission; therapy; diagnostics; agonists;  
 KW antagonists; cerebral ischemia; epilepsy; memory; learning.  
 XX  
 OS Mus sp.  
 XX  
 PN EP600278-A.  
 PN  
 PD 08-JUN-1994.  
 XX  
 PF 11-NOV-1993; 93EP-0118298.  
 XX  
 PR 13-NOV-1992; 92JP-0303878.  
 XX  
 PA (MITU ) MITSUBISHI KASEI CORP.  
 XX  
 PI Mishina M;  
 PI  
 DR WPI: 1994-177637/22.  
 DR N-PSDB; AA065668.  
 XX  
 PT Glutamate receptor protein - for screening an agonist or  
 PT antagonist which binds to the receptor  
 XX  
 PS Example 1; Page 4; 20pp; English.  
 XX  
 CC N-methyl-D-aspartic acid (NMDA) type glutamate receptors play a  
 CC central role in nervous information transmission, therefore the  
 CC clarification of the molecular structure and function will help  
 CC understanding of the mechanisms involved in the process. Highly  
 CC preserved amino acid sequences (AAR55249, and AAR55250 which exists at  
 CC the upstream end of the membrane spanning region of M3) of mouse  
 CC NMDA type receptor subunits were used to design primers (AA065667-68)  
 CC for use in cloning of an NMDA type glutamate receptor (AA065666 and  
 CC AAR55248). The NMDA type glutamate receptor is useful for screening  
 CC for an agonist or antagonist which binds the receptor. The gene is  
 CC useful for clarifying nervous information transmission at a synapse,  
 CC etc., but may also be useful for therapy of genetic diseases and  
 CC prepn. of novel pharmaceuticals.  
 XX  
 SQ Sequence 7 AA;  
 Query Match 100.0%; Score 18; DB 15; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 NLAA 4  
 Db 4 nlaa 7

```

RESULT 4
ID AAW23434 standard; peptide: 7 AA.
XX
AC AAW23434;
XX
DT 23-APR-1998 (first entry)
XX
DE CDR-2 of IW12 light chain.
XX
KW Antibody; complementarily determining region; CDR; light chain; rat; CEA;
KW IW12; anti-carcinoembryonic antigen; vaccine; immune response; cancer;
XX therapy; Idiotypic region.
XX
OS Rattus sp.
XX
PN WO9734636-A1.
XX
PD 25-SEP-1997.
XX
PF 19-MAR-1997; 97WO-US04696.
XX
PR 20-MAR-1996; 96US-0013708.
XX
PA (IMMU-) IMMUNOMEDICS INC.
XX
PI Hansen H, Leung S, Losman MJ;
XX
DR WPI; 1997-479997/44.
XX
PT Chimeric and humanised anti-carcinoembryonic antigen anti-idiotype
PT antibodies - useful in detection of anti-CEA antibodies and as
PT vaccine to stimulate immune response against cancer
XX
PS Claim 4; Page 30; 46pp; English.
XX
CC This sequence represents the complementarity determining region-2 (CDR-2)
CC of the IW12 light chain. This sequence is used in an antibody of the
CC invention. The antibody of the invention is a chimeric or humanised
CC anti-idiotype antibodies (cab and hab, respectively) or a fragment which
CC specifically binds to the idiotype region of an anti-carcinoembryonic
CC antigen (CEA), where: (i) cab comprises the IW12 light (L) and
CC heavy (H) chain variable regions, or silent mutations; and (ii) hab
CC comprises IW12 complementarity determining regions (CDR) and humanised
CC framework (FR) regions. The hab is used as a vaccine to stimulate an
CC immune response in a patient against cancers expressing CEA. The hab, can
CC be used to clear non-targeted antibody in a method of diagnosis or
CC treatment of a patient where a CEA antibody is used as a (pre-)targeting
CC or therapy agent. The cab is used to detect the presence of an antibody
CC that specifically binds to CEA in a sample.
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 18; DB 18; Length 7;
Best Local Similarity 100.0%; Pred. No. 3.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4
   ||||
DB 4 nlaa 7

RESULT 5
ID AAB6585 standard; Peptide: 7 AA.
XX
AC AAB6585;
XX
DE 12-APR-2001 (first entry)
XX

```

```

DE Human memapsin 2 inhibitor OM99-1.
XX
XX Human memapsin 2; neurotrophic; neuroprotective; amyloid precursor protein;
KW APP; memapsin 2 inhibitor; Alzheimer's disease.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 3..4 /note="residues 3 and 4 form a Leu-Ala dipeptide
FT isostere"
XX
XX WO200100665-A2.
XX
PD 04-JAN-2001.
XX
PF 27-JUN-2000; 2000WO-US17742.
XX
PR 28-JUN-1999; 99US-0141363.
XX
PR 30-NOV-1999; 99US-0168060.
XX
PR 25-JAN-2000; 2000US-0177836.
XX
PR 27-JAN-2000; 2000US-0178368.
XX
PR 08-JUN-2000; 2000US-0210292.
XX
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
XX
PA (UNIT ) UNIV ILLINOIS FOUND.
XX
PI Tang JUN, Hong L, Ghosh AK;
XX
DR WPI; 2001-137933/14.
XX
PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2
PT having 2 catalytic aspartic residues and substrate binding cleft, used
PT to treat Alzheimer's disease by blocking amyloid precursor protein
PT cleavage
XX
XX Example 7; Page 36; 86pp; English.
XX
PS The present sequence is given in a specification relating to an inhibitor
XX of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
XX active site, which is defined by the presence of two catalytic aspartic
XX residues and a substrate binding cleft. The inhibitor is useful for
XX the treatment and diagnosis of Alzheimer's disease. It is useful in
XX screens for individuals with a genetic predisposition to Alzheimer's
XX disease. The inhibitor is useful as a reagent for specifically binding to
XX memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
XX isolation, purification and characterisation.
XX
SQ Sequence 7 AA;

Query Match 100.0%; Score 18; DB 22; Length 7;
Best Local Similarity 100.0%; Pred. No. 3.4e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4
   ||||
DB 2 nlaa 5

RESULT 6
ID AAB61347 standard; peptide: 7 AA.
XX
AC AAB61347;
XX
DE 02-APR-2001 (first entry)
XX
DE Peptide OM99-1.
XX
KW Memapsin 2; catalyst; Alzheimer's.
XX
OS Unidentified.

```

XX WO200100663-A2.  
 PN 04-JAN-2001.  
 PD 27-JUN-2000; 2000WO-US17661.  
 PF 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA Tang JUN, Lin X, Koelsch G;  
 PI WPI; 2001-102885/11.  
 DR Purified recombinant catalytically active memapsin 2, used to screen  
 XX inhibitors of it, which are used to treat and prevent Alzheimer's  
 PT disease -  
 PS Example 7; Page 36; 86pp; English.  
 XX The present invention relates to a purified recombinant  
 CC catalytically active memapsin 2. The invention may be used for  
 CC isolating inhibitors which are used to treat or prevent  
 CC Alzheimer's disease. The invention may also be used to screen  
 CC for individuals more genetically prone to develop Alzheimer's  
 CC disease.  
 XX Sequence 7 AA:  
 SQ

Query Match 100.0%; Score 18; DB 22; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 ||||  
 Db 2 nlaa 5

RESULT 7  
 AAB6586  
 ID AAB6586 standard; Peptide: 8 AA.  
 AC AAB6586;  
 XX 12-APR-2001 (first entry)  
 DT Human memapsin 2 inhibitor OM99-2.  
 DE Human memapsin 2; nootropic; neuroprotective; amyloid precursor protein;  
 XX APP; memapsin 2 inhibitor; Alzheimer's disease.  
 KW Synthetic.  
 XX OS  
 XX Key Location/Qualifiers  
 FH Modified-site 4..5  
 FT /note="residues 4 and 5 form a Leu-Ala dipeptide  
 FT isostere"  
 FT  
 XX WO200100665-A2.  
 PN 04-JAN-2001.  
 PD 27-JUN-2000; 2000WO-US17742.  
 PF 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA (UNIT ) UNIV ILLINOIS FOUND.  
 XX Tang JUN, Hong L, Ghosh AK;  
 PI WPI; 2001-137933/14.  
 DR Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 XX having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage -  
 PS Example 7; Page 36; 86pp; English.  
 XX The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 XX Sequence 8 AA:  
 SQ

Query Match 100.0%; Score 18; DB 22; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 ||||  
 Db 3 nlaa 6

RESULT 8  
 AAB61348  
 ID AAB61348 standard; peptide: 8 AA.  
 AC AAB61348;  
 XX 02-APR-2001 (first entry)  
 DT Peptide OM99-2.  
 DE Memapsin 2; catalyst; Alzheimer's.  
 XX KW  
 XX Unidentified.  
 XX OS  
 XX WO200100663-A2.  
 PN 04-JAN-2001.  
 PD 27-JUN-2000; 2000WO-US17661.  
 PF 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA Tang JUN, Lin X, Koelsch G;  
 PI WPI; 2001-102885/11.  
 DR Purified recombinant catalytically active memapsin 2, used to screen  
 XX inhibitors of it, which are used to treat and prevent Alzheimer's  
 PT

PT disease -  
XX  
PS Example 7; Page 36; 86pp; English.  
XX  
CC The present invention relates to a purified recombinant  
CC catalytically active memapsin 2. The invention may be used for  
CC isolating inhibitors which are used to treat or prevent  
CC Alzheimer's disease. The invention may also be used to screen  
CC for individuals more genetically prone to develop Alzheimer's  
CC disease.  
XX  
SQ Sequence 8 AA:  
  
Query Match 100.0%; Score 18; DB 22; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 NLAA 4  
   1111  
DB 3 nlaa 6  
  
RESULT 9  
AAV46543  
ID AAV46543 standard; Peptide: 9 AA.  
XX  
AC AAV46543;  
XX  
DT 01-DEC-1999 (first entry)  
XX  
DE Immunogenic peptide having a human leukocyte antigen binding motif #1154.  
XX  
KW Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;  
KW immune response; T cell activation; major histocompatibility complex;  
KW cytotoxic T lymphocytes; CTL; tumour rejection; viral infection; cancer;  
KW prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;  
KW vaccine; immunisation.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO9455954-A1.  
XX  
PD 16-SEP-1999.  
XX  
PF 13-MAR-1998; 98WO-US05039.  
XX  
PR 13-MAR-1998; 98WO-US05039.  
XX  
PA (EPTM-) EPTMUNE INC.  
XX  
PI Sette A, Kubo RT, Sidney J, Cells E, Grey HM, Southwood S;  
XX  
DR WPI: 1999-551214/46.  
XX  
PT New immunogenic peptides with HLA binding motif, useful in treatment  
PT and diagnosis of cancers and viral diseases -  
XX  
PS Claim 1; Page 75; 150pp; English.  
XX  
XX AAV45390 to AAV48214 represent specifically claimed immunogenic peptides  
CC having a human major histocompatibility complex (MHC) Class I (also  
CC known as human leukocyte antigen (HLA)) binding motif. The immunogenic  
CC peptides can bind to a specific HLA allele (i.e. HLA-A subtypes  
CC HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell  
CC response against the antigen from which the peptide is derived.  
CC Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are  
CC normally induced by an antigen in the form of a peptide fragment bound  
CC to a HLA molecule, rather than the intact foreign antigen itself, and  
CC are particularly important in tumour rejection and in fighting viral  
CC infections. The peptides are therefore useful therapeutically to treat  
CC or prevent viral infections and cancers in mammals (especially humans)

CC e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.  
CC They can be administered as vaccines to elicit an immune response in  
CC individuals susceptible or otherwise at risk of viral infection or  
CC cancer, or used to treat chronic or acute conditions. They are also  
CC useful diagnostically, and can be used to induce a cytotoxic T cell  
CC response, by contacting a cytotoxic T cell with the peptide e.g. to  
CC produce CTLs ex vivo for infusion back into a patient. The  
CC polynucleotides encoding the immunogenic peptides are also useful  
CC therapeutically and for immunisation as above.  
XX  
SQ Sequence 9 AA:  
  
Query Match 100.0%; Score 18; DB 20; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 NLAA 4  
   1111  
DB 4 nlaa 7  
  
RESULT 10  
AAB07879  
ID AAB07879 standard; peptide: 9 AA.  
XX  
AC AAB07879;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE A peptide fragment derived from beta-amyloid precursor protein.  
XX  
KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.  
XX  
OS Homo sapiens.  
XX  
PN WO200047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PF 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
XX  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basl G, Doane MT, Frigon N, John V, Power M,  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
DR WPI: 2000-533011/48.  
XX  
PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.  
PT Alzheimer's disease -  
XX  
PS Disclosure; Page 12; 121pp; English.  
XX  
XX The specification describes a beta-secretase enzyme. The enzyme cleaves  
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a peptide derived from beta-amyloid  
CC precursor protein  
XX  
SQ Sequence 9 AA:

Query Match 100.0%; Score 18; DB 21; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 ||||  
 DB 4 nlaa 7

## RESULT 11

AAV77829  
 ID AAV77829 standard; peptide; 10 AA.

AC AAV77829;

DT 06-JUN-2000 (first entry)

DE Human chromogranin B protein granin region.

KW BRCA; tumour; prostate cancer; cytostatic; antiproliferative;

XX gene therapy; granin domain; chromogranin B.

OS Homo sapiens.

PN CA2218197-A1.

PD 05-JUN-1999.

PF 12-DEC-1997; 97CA-2218197.

PR 05-DEC-1997; 97US-0986106.

PA (UNIW ) UNIV WASHINGTON.

XX (UYVA-) UNIV VANDERBILT.

PI Robinson-Benion CL, Thompson ME, Holt JT, Jensen RA, Steiner MS;

XX King M;

DR WPI: 2000-238071/21.

PT New method of treatment and suppression of prostate cancer comprises  
 PT using the BRCA family of genes to decrease the growth rate of the tumor  
 PT

PS Example 4; Fig 5; 166pp; English.

XX The invention relates to a method for suppressing the growth of a  
 CC prostate tumour in a mammal that comprises introducing to the tumour a  
 CC vector comprising a nucleic acid sequence encoding a BRCA family gene  
 CC product operatively linked to a promoter, where production of the BRCA  
 CC family gene product results in a decrease in the growth rate of the  
 CC tumour. The methods are used to suppress the growth of and also to treat  
 CC prostate tumour in a mammal where the tumour is gene-linked hereditary  
 CC represent granin regions of selected granin family members.

XX Sequence 10 AA;

Query Match 100.0%; Score 18; DB 21; Length 10;

Best Local Similarity 100.0%; Pred. No. 68;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 ||||  
 DB 2 nlaa 5

## RESULT 12

AAV77830  
 ID AAV77830 standard; peptide; 10 AA.

XX AAV77830;

XX 06-JUN-2000 (first entry)

DT Bovine chromogranin B protein granin region.

KW BRCA; tumour; prostate cancer; cytostatic; antiproliferative;  
 XX gene therapy; granin domain; chromogranin B.

OS Bos sp.

PN CA2218197-A1.

PD 05-JUN-1999.

PF 12-DEC-1997; 97CA-2218197.

PR 05-DEC-1997; 97US-0986106.

PA (UNIW ) UNIV WASHINGTON.

XX (UYVA-) UNIV VANDERBILT.

PI Robinson-Benion CL, Thompson ME, Holt JT, Jensen RA, Steiner MS;

XX King M;

DR WPI: 2000-238071/21.

PT New method of treatment and suppression of prostate cancer comprises  
 PT using the BRCA family of genes to decrease the growth rate of the tumor  
 PT

PS Example 4; Fig 5; 166pp; English.

XX The invention relates to a method for suppressing the growth of a  
 CC prostate tumour in a mammal that comprises introducing to the tumour a  
 CC vector comprising a nucleic acid sequence encoding a BRCA family gene  
 CC product operatively linked to a promoter, where production of the BRCA  
 CC family gene product results in a decrease in the growth rate of the  
 CC tumour. The methods are used to suppress the growth of and also to treat  
 CC prostate tumour in a mammal where the tumour is gene-linked hereditary  
 CC represent granin regions of selected granin family members.

XX Sequence 10 AA;

Query Match 100.0%; Score 18; DB 21; Length 10;

Best Local Similarity 100.0%; Pred. No. 68;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 ||||  
 DB 2 nlaa 5

## RESULT 13

AAV77831  
 ID AAV77831 standard; peptide; 10 AA.

AC AAV77831;

DT 06-JUN-2000 (first entry)

DE Mouse chromogranin B protein granin region.

KW BRCA; tumour; prostate cancer; cytostatic; antiproliferative;

XX gene therapy; granin domain; chromogranin B.

OS Mus sp.

PN CA2218197-A1.

PD 05-JUN-1999.  
 XX  
 PF 12-DEC-1997; 97CA-2218197.  
 XX  
 PR 05-DEC-1997; 97US-0986106.  
 XX  
 PA (UNITM ) UNIV WASHINGTON.  
 PA (UYVA-) UNIV VANDERBILT.  
 PI Robinson-Benton CL, Thompson ME, Holt JT, Jensen RA, Steiner MS;  
 PI King M;  
 PI WPI: 2000-238071/21.  
 DR  
 XX  
 PT New method of treatment and suppression of prostate cancer comprises  
 PT using the BRCA family of genes to decrease the growth rate of the tumor  
 PT -  
 PS  
 XX Example 4: Fig 5; 166pp: English.  
 XX  
 CC The invention relates to a method for suppressing the growth of a  
 CC prostate tumour in a mammal that comprises introducing to the tumour a  
 CC vector comprising a nucleic acid sequence encoding a BRCA family gene  
 CC product operatively linked to a promoter, where production of the BRCA  
 CC family gene product results in a decrease in the growth rate of the  
 CC tumour. The methods are used to suppress the growth of and also to treat  
 CC prostate tumour in a mammal where the tumour is gene-linked hereditary  
 CC prostate cancer or sporadic prostate cancer. Sequences AAV7821-841  
 CC represent granlin regions of selected granlin family members.  
 XX  
 SQ Sequence 10 AA:

Query Match 100.0%; Score 18; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 IIII  
 DB 2 nlaa 5

RESULT 14  
 AAR89266  
 ID AAR89266 standard; Peptide: 12 AA.  
 XX  
 AC AAR89266;  
 XX  
 DT 07-APR-1996 (first entry)  
 XX  
 DE Heparinase-II-derived peptide-2A.  
 XX  
 KW Heparinase-II; heparin degradation; heparan sulphate degradation;  
 KW Escherichia coli; polymerase chain reaction; PCR; primer; toxicity;  
 KW cloning; vector; ribosome binding site; haemostatic; blood-clotting;  
 KW antibody; affinity chromatography.  
 XX  
 OS Flavobacterium heparinum.  
 XX  
 PN WO9534635-A1.  
 XX  
 PD 21-DEC-1995.  
 XX  
 PF 09-JUN-1995; 95WO-US07391.  
 XX  
 PR 10-JUN-1994; 94US-0258639.  
 XX  
 PA (IBEX-) IBEX TECHNOLOGIES.  
 PA (ZIMM/) ZIMMERMANN J.  
 XX  
 PI Bennett C, Blain F, Gu K, Musil R, Su H, Zimmermann J;  
 XX  
 DR WPI: 1996-097381/10.

XX  
 PT Nucleic acids encoding Flavobacterium heparinum heparinase II and  
 PT III - for degrading heparin and heparan sulphate, also related host  
 XX cells, proteins and antibodies, useful in heparinase purificn.  
 XX  
 PS Example 2: Page 21: 75pp: English.  
 XX

CC The sequence corresponds to peptide-2A derived from Flavobacterium  
 CC heparinum heparinase-II (AAR89264), which degrades heparin and heparan  
 CC sulphate. The sequence, along with peptides 2B (AAR89267) and 2C  
 CC (AAR89268) may be used to construct DNA primers, e.g. 2-1 to 2-4  
 CC (AAQ99228-099231), which may be used to isolate the heparinase-II  
 CC gene from a gene library in phage lambda-DASH-II in Escherichia  
 CC coli by polymerase chain reaction amplification. Toxicity of the  
 CC gene and natural selection of the host against clones with the  
 CC entire sequence has been circumvented by cloning sections of the  
 CC gene separately in a vector with a modified ribosome binding site,  
 CC which increases expression levels. The heparinase-II may be used  
 CC to neutralise anticoagulant activity. Antibodies against the  
 CC protein may be used to differentiate between native and  
 CC recombinant enzymes, and when immobilised they may be used for  
 CC heparinase purification by affinity chromatography.  
 XX  
 SQ Sequence 12 AA:

Query Match 100.0%; Score 18; DB 17; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLAA 4  
 IIII  
 DB 7 nlaa 10

RESULT 15  
 AAR32954  
 ID AAR32954 standard; peptide: 14 AA.  
 XX  
 AC AAR32954;  
 XX  
 DT 02-JUL-1993 (first entry)  
 XX  
 DE Mastoparan analogue (17) having G protein modulatory activity.  
 XX  
 KW Mastoparan; MP: cellular regulation; receptor-based analogue;  
 KW G protein; guanine nucleotide binding regulatory protein;  
 KW asthma; ulcer; cardiovascular; disease; allergy; Parkinson's;  
 KW small cell carcinoma; lung; glaucoma; respiratory tract congestion;  
 KW inflammation.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9303749-A.  
 XX  
 PD 04-MAR-1993.  
 XX  
 PF 14-AUG-1992; 92WO-US06825.  
 XX  
 PR 21-AUG-1991; 91US-0748319.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Hlgashijima T, Ross EM;  
 PI WPI: 1993-093715/11.  
 DR  
 XX  
 PT New peptide(s) contg. mastoparan- or receptor-analogue region -  
 PT uses as G protein modulators, for treating asthma, ulcers,  
 PT cardiovascular disorders and Parkinson's disease  
 XX  
 PS Claim 19; Page 80; 96pp: English.  
 XX

CC The peptide is an example of a highly generic formula, and is  
 CC represented as found in the disclosure of the specification. The  
 CC claimed peptide lacks the N-terminal IN amino acids.  
 CC The peptide is capable of modulating G protein action in a cell. It  
 CC may therefore be used for treating diseases involving G proteins,  
 CC e.g. asthma, ulcers, cardiovascular diseases, allergies, Parkinson's  
 CC disease, small cell carcinoma of the lung, glaucoma, respiratory  
 CC tract congestion or inflammation.

XX  
 SQ Sequence 14 AA;

Query Match 100.0%; Score 18; DB 14; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLAA 4  
 | | | |  
 DB 2 nlaa 5

Search completed: September 6, 2001, 16:43:32  
 Job time: 361 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:56 : Search time 231.42 Seconds  
(without alignments)  
4.574 Million cell updates/sec

Title: US-09-603-713-28  
Perfect score: 38  
Sequence: 1 EVNLAEF 8

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_16:\*

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.mhc:\*
- 8: sp.organelle:\*
- 9: sp.phage:\*
- 10: sp.plant:\*
- 11: sp.rodent:\*
- 12: sp.unclassified:\*
- 13: sp.vertebrate:\*
- 14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	32	84.2	123	10	Q9FKH5 arabidopsis
2	32	84.2	538	10	Q9FZHI arabidopsis
3	31	81.6	466	10	Q9ZUL1 arabidopsis
4	31	81.6	2270	14	Q9JFN3 arabidopsis
5	30	78.9	245	2	Q9XDB1 arabidopsis
6	30	78.9	280	1	Q9HMI8 arabidopsis
7	30	78.9	283	10	Q48807 arabidopsis
8	30	78.9	295	10	Q65376 arabidopsis
9	30	78.9	459	2	Q9PNAS arabidopsis
10	30	78.9	602	2	Q9F165 arabidopsis
11	29	76.3	153	2	Q9EYX8 arabidopsis
12	29	76.3	282	2	Q30485 arabidopsis
13	29	76.3	328	2	Q66974 arabidopsis
14	29	76.3	335	3	Q9P6L3 arabidopsis
15	29	76.3	416	2	Q57456 arabidopsis
16	29	76.3	529	14	Q39281 arabidopsis
17	29	76.3	643	2	Q9Z6G3 arabidopsis
18	29	76.3	657	3	Q9U0U3 arabidopsis
19	29	76.3	738	2	Q59490 arabidopsis

20	29	76.3	797	5	Q9GV13 hydra magni
21	28	73.7	68	1	Q9HQL1 halobacteri
22	28	73.7	177	10	Q9FXZ6 glycine max
23	28	73.7	237	2	Q05509 bacillus su
24	28	73.7	243	5	Q22341 caenorhabd
25	28	73.7	245	2	Q9KJVB rhizobium l
26	28	73.7	245	10	Q9LDV6 oryza sativ
27	28	73.7	253	2	Q48968 mycoplasma
28	28	73.7	293	4	Q60812 homo sapien
29	28	73.7	295	10	Q9SHF7 arabidopsis
30	28	73.7	306	6	Q77768 oryctolagus
31	28	73.7	313	11	Q9Z204 mus musculu
32	28	73.7	315	2	P94419 bacillus su
33	28	73.7	328	5	Q44501 caenorhabd
34	28	73.7	381	2	Q69324 pseudomonas
35	28	73.7	400	4	Q00328 homo sapien
36	28	73.7	494	11	Q9JLJ3 ratu
37	28	73.7	494	11	Q9JLJ2 mus musculu
38	28	73.7	587	5	Q76753 entamoeba h
39	28	73.7	600	2	Q9KG45 bacillus ha
40	28	73.7	629	3	Q9P4E1 cunn
41	28	73.7	699	2	Q51287 neisseria m
42	28	73.7	754	13	P79816 oryza sat
43	28	73.7	888	3	Q9HEP0 candida alb
44	28	73.7	906	5	Q17281 botryllus s
45	28	73.7	942	14	Q96585 avian adeno

## ALIGNMENTS

RESULT 1

Q9FKH5 PRELIMINARY: PRT: 123 AA.

Q9FKH5: 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE SIMILARITY TO TAIL-LIKE NON-ITR RETROELEMENT PROTEIN.

OS Arabidopsis thaliana (Mouse-ear cress). Tracheophyta; Spermatophyta; OC Eukaryota; Viridiplantae; Embryophyta; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_Taxid=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RX MEDLINE=98344145; PubMed=9679202;

RA Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence features of the regions of 1,381,565 bp covered by twenty one physically assigned P1 and TAC clones.";

RT DNA Res. 5:131-145(1998).

DR EMBL: AB011485; BAB09272.1; "

SQ SEQUENCE: 123 AA; 14378 MW; C0C734B59CA9553F CRC64;

Query Match 84.2%; Score 32; DB 10; Length 123;  
Best Local Similarity 87.5%; Pred. No. 8.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVNLAEF 8

DB 110 EVNLAEF 117

RESULT 2

Q9FZHI PRELIMINARY: PRT: 538 AA.

Q9FZHI: 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE F5A13.3 PROTEIN.  
 GN F5A13.3.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
 RA Altieri H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,  
 RA Buehler E., Chao Q., Chin C., Chio J., Choi E., Gonzalez A.,  
 RA Hong B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,  
 RA Lenz C., Liu A., Liu S., Mukharsy N., Pham P., Sakano H., Shin P.,  
 RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.,  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC008046; AAF9831.1; --  
 SQ SEQUENCE 538 AA; 59440 MW; 99C79074D5EFA315 CRC64;

Query Match 84.2%; Score 32; DB 10; Length 538;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYNLAEE 7  
 DB 532 EYNLAEE 538  
 |||||

RESULT 3  
 O92U11 PRELIMINARY; PRT; 466 AA.  
 ID O92U11  
 AC O92U11  
 DT 01-MAY-1999 (TReMBLrel. 10, Created)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)  
 DE F27D4.1 PROTEIN.  
 GN F27D4.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-CV. COLUMBIA;  
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,  
 RA Shen M., Rouning C.M., Fraser C.M., Somerville C.R., Venter J.C.,  
 RT "Arabidopsis thaliana chromosome II BAC F27D4 genomic sequence."  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC005967; AAD03372.1; --  
 SQ SEQUENCE 466 AA; 52211 MW; EDE1AA7D77FE0A57 CRC64;

Query Match 81.6%; Score 31; DB 10; Length 466;  
 Best Local Similarity 75.0%; Pred. No. 63;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYNLAEE 8  
 DB 252 EYNLAEE 259  
 |||||

RESULT 4  
 O9JFN3 PRELIMINARY; PRT; 2270 AA.  
 ID O9JFN3  
 AC O9JFN3  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE RNA POLYMERASE.  
 RL RMA POLYMERASE.  
 GN L.

OS Tupala paramyxovirus (TPMV).  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Paramyxovirinae; Paramyxovirus.  
 OX NCBI\_TaxID=92129;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tidona C.A., Daral G.,  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF079780; AAF63393.1; --  
 DR InterPro: IPR001016; --  
 DR Pfam: PF00946; Paramyx\_RNA-pol; 1.  
 SQ SEQUENCE 2270 AA; 259395 MW; 037C728466E95D95 CRC64;

Query Match 81.6%; Score 31; DB 14; Length 2270;  
 Best Local Similarity 75.0%; Pred. No. 36+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYNLAEE 8  
 DB 350 DVNLTAEE 357  
 |||||

RESULT 5  
 O9XDB1 PRELIMINARY; PRT; 245 AA.  
 ID O9XDB1  
 AC O9XDB1  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE REGULATORY PROTEIN.  
 GN PIPR.  
 OS Mycobacterium smegmatis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterinae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1772;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-MC2155.  
 RC MEDLINE-99287823; PubMed-10348853;  
 RX Poupin P., Ducrocq V., Hallier-Soulier S., Truffaut N.;  
 RA "Cloning and characterization of the genes encoding a cytochrome P450  
 (P450) involved in piperidine and pyroliidine utilization and its  
 RT regulatory protein (Pipr) in Mycobacterium smegmatis mc2155."  
 RL J. Bacteriol. 181:3419-3426(1999).  
 DR EMBL: AF102509; AAD28343.1; --  
 DR InterPro: IPR000524; --  
 DR SMART: PF00392; gntR; 1.  
 DR SMART: SM00345; HTH\_GNTR; 1.  
 SQ SEQUENCE 245 AA; 26363 MW; 05BCBF0F3EC27DF9 CRC64;

Query Match 78.9%; Score 30; DB 2; Length 245;  
 Best Local Similarity 75.0%; Pred. No. 54;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYNLAEE 8  
 DB 41 EVDLAEE 48  
 |||||

RESULT 6  
 O9HNT8 PRELIMINARY; PRT; 280 AA.  
 ID O9HNT8  
 AC O9HNT8  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE PHOSPHONATES TRANSPORT ATP-BINDING.  
 GN PHN OR VNG2085G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Halobacterium; Halobacteriales; Halobacteriaceae;  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacterium.

OX NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasry S.R., Baliga N.S., Thorsson V., Sirogna J.,  
 RA Swartzell S., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Leithauer B., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Madocks D.G., Peck R.F., Pohlischer M., Spidich J.L., Jung K.-H.,  
 RA Isebnager T.A., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Alam M., Freitas T., Liang P., Riley M., Hood L., Dassarma S.;  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RT Genome sequence of *Halobacterium* species NRC-1.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL: AE005098; AAG20232.1; -;  
 DR InterPro: IPR001687; -;  
 DR InterPro: IPR002078; -;  
 DR InterPro: IPR003439; -;  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 DR ATP-binding.  
 KM  
 SQ SEQUENCE 280 AA; 30083 MW; E147D27A4985FA88 CRC64;

Query Match 78.9%; Score 30; DB 1; Length 280;  
 Best Local Similarity 75.0%; Pred. No. 62;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 Db 198 QVNLAAEF 205

RESULT 7  
 ID 048807 PRELIMINARY; PRT; 283 AA.  
 AC 048807;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE F2401.16.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chlou J., Choi E.,  
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Theologis A., Tortum L., Vaysberg M., Yu G., Davis R., Federspiel N.,  
 RA Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.  
 RL EMBL: AC003113; AAF70846.1; -;  
 DR InterPro: IPR001599; -;  
 DR PROSITE: PS00477; ALPHA\_2\_MACROGLOBULIN; UNKNOWN\_1.  
 DR PROSITE: PS00477; ALPHA\_2\_MACROGLOBULIN; UNKNOWN\_1.  
 SQ SEQUENCE 283 AA; 32218 MW; 2B805E9DB92C52B9 CRC64;

Query Match 78.9%; Score 30; DB 10; Length 283;  
 Best Local Similarity 75.0%; Pred. No. 63;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 Db 159 ETLNAGEF 166

RESULT 8

065376  
 ID 065376 PRELIMINARY; PRT; 295 AA.  
 AC 065376;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)  
 DE F12F1.10 PROTEIN.  
 GN F12F1.10.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Vysotskaia V.S., Osborne B.I., Schwartz J.R., Tortum M., Yu G.,  
 RA Kwan A., Oll O., Liu S., Buehler E., Conway A.B., Conway A.R.,  
 RA Dewar K., Feng J., Kim C., Kurtz D., Palm C.J., Li Y., Shinn P.,  
 RA Sun H., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;  
 RA Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AC002131; AAC17612.1; -;  
 KM  
 SQ SEQUENCE 295 AA; 33658 MW; 9A11A6CBF1C14E45 CRC64;

Query Match 78.9%; Score 30; DB 10; Length 295;  
 Best Local Similarity 75.0%; Pred. No. 66;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 Db 159 ETLNAGEF 166

RESULT 9  
 ID 09PNA5 PRELIMINARY; PRT; 459 AA.  
 AC 09PNA5;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE PUTATIVE MCP-DOMAIN SIGNAL TRANSDUCTION PROTEIN.  
 GN C01190C.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,  
 RA Tagatz K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,  
 RA Whitehead S., Barrett B.G.;  
 RT "The genome sequence of the food-borne pathogen *Campylobacter jejuni*  
 RT reveals hypervariable sequences.";  
 RL Nature 403:665-668(2000).  
 DR EMBL: AL139077; CAB73444.1; -;  
 DR InterPro: IPR000122; -;  
 DR Pfam: PF000015; MCP\_Signal; 1.  
 DR SMART: SM00283; MA; 1.  
 SQ SEQUENCE 459 AA; 51016 MW; DDBCCF54EFA96DFE CRC64;

Query Match 78.9%; Score 30; DB 2; Length 459;  
 Best Local Similarity 75.0%; Pred. No. 1; Indels 0; Gaps 0;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 Db 407 ETLNAGEF 414

RESULT 10  
 Q9P165 PRELIMINARY; PRT; 602 AA.  
 AC Q9P165: 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE METHYLMAIONYL-COA MUTASE SMALL SUBUNIT (EC 5.4.99.2).  
 GN CMCA.  
 OS Amycolatopsis mediterranei.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae; Amycolatopsis.  
 NC NCB1\_TaxID=33910;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=U-32;  
 RA Zhang W.W., Yang L., Jiang W.H., Chiao J.S., Zhao G.P.;  
 RT "Cloning, sequencing and expression of the gene encoding  
 RT methylmaionyl-coa mutase from Amycolatopsis mediterranei";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF117980; AAC47627.1; -;  
 KW Isomerase.  
 SQ SEQUENCE 602 AA; 62668 MW; D396FD3D76353A7 CRC64;

Query Match 78.9%; Score 30; DB 2; Length 602;  
 Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVNLAEF 8  
 ID 1:|||||  
 Db 5 ELDLAEF 12

RESULT 11  
 Q9EVX8 PRELIMINARY; PRT; 153 AA.  
 ID Q9EVX8: 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE TRANSKETOLASE (FRAGMENT).  
 GN RECP.  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 NC NCB1\_TaxID=1314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Enright M.C., Spratt B.G., Kalta A., Bessen D.E.;  
 RT "Multilocus sequence typing of Streptococcus pyogenes and the  
 RT relationships between emm type and clone."  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ407042; CAC21161.1; -;  
 FT NON\_TER. 1  
 FT NON\_TER. 153  
 FT SEQUENCE 153 AA; 16659 MW; 396C4C1120488E21 CRC64;

Query Match 76.3%; Score 29; DB 2; Length 153;  
 Best Local Similarity 75.0%; Pred. No. 1.5;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAEF 8  
 ID 1:|||||  
 Db 97 EHLIAEF 104

RESULT 12  
 Q30485 PRELIMINARY; PRT; 282 AA.

AC Q30485: 01-JAN-1998 (TREMblrel. 05, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)  
 DE HYPOTHEICAL 29.8 KDA PROTEIN.  
 OS Streptomyces hygroscopicus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 NC NCB1\_TaxID=1912;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29253;  
 RA MEDLINE=98085969; PubMed=9426000;  
 RA Ruan X., Stassi D., Lax S., Katz L.;  
 RT "A second type-I PKS gene cluster isolated from Streptomyces  
 RT hygroscopicus ATCC 29253, a rapamycin-producing strain."  
 RL Gene 203:1-9(1997).  
 DR EMBL: AF007101; AAC38067.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 282 AA; 29831 MW; 6D9B89C3A205F8F CRC64;

Query Match 76.3%; Score 29; DB 2; Length 282;  
 Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVNLAEF 7  
 ID 1:|||||  
 Db 111 QVNLAEF 117

RESULT 13  
 Q66974 PRELIMINARY; PRT; 328 AA.  
 ID Q66974: 01-AUG-1998 (TREMblrel. 07, Created)  
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
 DE HYPOTHEICAL 36.0 KDA PROTEIN.  
 GN AQ\_778.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
 NC NCB1\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RX Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 RT aeolicus."  
 RL Nature 392:353-358(1998).  
 DR EMBL: AE000707; AAC06942.1; -;  
 DR InterPro: IPR000873; -;  
 DR InterPro: IPR002882; -;  
 DR Pfam: PF01933; UPR0052; 1.  
 DR PROSITE: PS00455; AMP\_BINDING; UNKNOWN\_1.  
 DR Hypothetical protein.  
 KW Hypothetical protein.  
 SQ SEQUENCE 328 AA; 35957 MW; 30BB214E9958F6B CRC64;

Query Match 76.3%; Score 29; DB 2; Length 328;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 VNLADEF 8  
 ID 1:|||||  
 Db 133 VNLADEF 139

RESULT 14  
 Q9P6L3

Search completed: September 6, 2001, 16:49:56  
Job time: 740 sec

ID 09PEL3 PRELIMINARY; PRT; 335 AA.  
AC 09PEL3;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE CUP9 INTERACTING PROTEIN SCNIP.  
GN SCNIP.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972H-;  
RA Aert R., Robben J., Volckaert G., Wood V., Rajandream M.A.,  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL355632; CAB90779.1; -  
SQ SEQUENCE 335 AA; 38459 MW; DIC41E4A9DA19CEF CRC64;

Query Match 76.3%; Score 29; DB 3; Length 335;  
Best Local Similarity 73.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
: | | | | |  
DB 171 QVRLAAEF 178

RESULT 15  
O57456 PRELIMINARY; PRT; 416 AA.  
AC 057456;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE HYPOTHETICAL 46.8 KDA PROTEIN.  
GN SLR0351.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96127529; PubMed=8590279;  
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
RA Sugiyura M., Tabata S.;  
RT \*Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
RT region from map positions 64% to 92% of the genome.";  
RL DNA Res. 2:153-166(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Cassier-Chauvat C., Poncellet M., Villioing S., Chauvat F.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: D64000; BAA10162.1; -  
DR EMBL: U38915; AAB72122.1; -  
DR InterPro: IPR001091; -  
DR PROSITE: PS00093; N4\_MTASE; UNKNOWN\_1.  
KW Hypothetical protein.  
SQ SEQUENCE 416 AA; 46828 MW; 43DA24914AA30D61 CRC64;

Query Match 76.3%; Score 29; DB 2; Length 416;  
Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVNLAAE 7  
: | | | | |  
DB 245 EVNLAAQ 251

Fri Sep 7 10:58:14 2001

us-09-603-713-28.rspt

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:11 ; Search time 72.75 Seconds  
(without alignments)  
3.767 Million cell updates/sec

Title: US-09-603-713-28

Perfect score: 38

Sequence: 1 EVNLAAEF 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	31	81.6	236	1	P47625 mycoplasma
2	30	78.9	2493	1	P33194 saccharomyc
3	29	76.3	387	1	P41890 sciliosacch
4	29	76.3	430	1	P73456 synechocyst
5	29	76.3	474	1	P08401 escherichia
6	29	76.3	509	1	P49819 canis famli
7	29	76.3	509	1	P09622 homo sapien
8	29	76.3	509	1	P03623 sus scrofa
9	29	76.3	614	1	O86781 s glucosami
10	29	76.3	623	1	GIMS_MYCTU
11	29	76.3	624	1	GIMS_MYCTE
12	29	76.3	627	1	GIMS_MYCSM
13	28	73.7	47	1	P11112 bacterioph
14	28	73.7	68	1	P17104 halobacteri
15	28	73.7	282	1	ROC_XENLA
16	28	73.7	289	1	P19600 xenopus lae
17	28	73.7	303	1	P21903 proionigen
18	28	73.7	463	1	P07910 homo sapien
19	28	73.7	509	1	P15628 soybean chl
20	28	73.7	509	1	O08749 mus musculu
21	28	73.7	676	1	RNR_CHLPN
22	28	73.7	734	1	RELA_BACSU
23	28	73.7	942	1	HEX_ADEGI
24	28	73.7	1131	1	VAB9_YEAST
25	28	73.7	1238	1	DROG_MOUSE
26	27	71.1	144	1	URE2_YERPS
27	27	71.1	160	1	MOAC_ECOLI
28	27	71.1	215	1	UCRI_YEAST
29	27	71.1	222	1	YFBI_ECOLI
30	27	71.1	232	1	YHUY_ECOLI
31	27	71.1	244	1	NOG4_RHIME
32	27	71.1	245	1	NODG_RHIME
33	27	71.1	245	1	NODG_RHIS3
					P73332 rhizobium s
					O57865 pyrococcus

34	27	71.1	257	1	YW64_PYRAB	O9v2f1 pyrococcus
35	27	71.1	262	1	ORC6_MOUSE	O9wuj8 mus musculu
36	27	71.1	285	1	AMPR_CITDI	P52658 citrobacter
37	27	71.1	292	1	AMPR_PROCT	O69772 providencia
38	27	71.1	295	1	NMCR_ENTCT	P52676 enterobacte
39	27	71.1	303	1	LMBI_CHICK	O01635 gallus galli
40	27	71.1	313	1	CBBR_RHIME	P36885 rhizobium m
41	27	71.1	352	1	KE4_BRARE	O9pub8 brachydanio
42	27	71.1	397	1	NIFS_KLEPN	P05344 klebsiella
43	27	71.1	399	1	PNCB_ECOLI	P18133 escherichia
44	27	71.1	428	1	FTSZ_ANASP	P45482 anabena sp
45	27	71.1	430	1	MENF_HAETN	P44613 haemophilus

## ALIGNMENTS

```
RESULT 1
ID Y385_MYCCE STANDARD: PRT: 236 AA.
AC P47625; Q49354;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MG385.
GN MG385.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Frieser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann J.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrman J.B.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucher T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 3-78 AND 87-176 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -1- SIMILARITY: TO GLYCEROPHOSPHORYL DIESTER PHOSPHODIESTERASES (EC
CC 3.1.4.46).
CC -1- SIMILARITY: M.GENTRALIUM MG293.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39720; AAC71612.1; -
DR EMBL: U02112; AAD12385.1; -
DR EMBL: U02246; AAA03402.1; -
DR TIGR: MG385; -
KW Hypothetical protein; Hydrolyase.
FT CONFLICT 87 S -> M (IN REF. 2).
SQ SEQUENCE 236 AA; 27733 MW; D7A3AF4E3630BDE2 CRC64;
```

Query Match 81.6%; Score 31; DB 1; Length 236;  
Best Local Similarity 75.0%; Pred. No. 4.4;





CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D90906; BAA17496.1; -  
DR HSSP; Q57816; 1FS2.  
DR InterPro; IPR000158; -  
DR InterPro; IPR003008; -  
DR Pfam; PF00091; tubulin; 1.  
DR PRINTS; PR00423; CELLDV1SFPSZ.  
DR PROSITE; PS01134; FTSZ\_1; 1.  
DR PROSITE; PS01135; FTSZ\_2; 1.  
KW Cell division; Septation; GTP-binding.  
FT NP\_BIND 159 167 GTP (POTENTIAL).  
SO SEQUENCE 430 AA; 44773 MW; E3C7DD534983FF44 CRC64;

Query Match 76.3%; Score 29; DB 1; Length 430;  
Best Local Similarity 85.7%; Pred. No. 26;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVNLAEE 7  
|||  
DB 329 EVNVAEE 335

RESULT 5  
CREC\_ECOLI STANDARD: PRT; 474 AA.  
ID CREC\_ECOLI  
AC P08401;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE SENSOR PROTEIN CREC (EC 2.7.3.-).  
GN CREC OR PHOM.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-87008393; PubMed-3531171;  
RA Amemura M., Makino K., Shinagawa H., Nakata A.;  
RT "Nucleotide sequence of the phom region of Escherichia coli: four  
RT open reading frames may constitute an operon.";  
RL J. Bacteriol. 168:294-302(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE-95334362; PubMed-7610040;  
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
RA Blattner F.R.;  
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
RT region from 92.8 through 100 minutes.";  
RL Nucleic Acids Res. 23:2105-2119(1995).  
RN [3]  
RP SEQUENCE OF 381-474 FROM N.A.  
RC STRAIN-K12;  
RX MEDLINE-88216172; PubMed-2835585;  
RA Drury L.S., Buxton R.S.;  
RT "Identification and sequencing of the Escherichia coli cel gene which  
RT codes for an inner membrane protein, mutation of which causes  
RT tolerance to colicin E2.";  
RL Mol. Microbiol. 2:109-119(1988).  
RN [4]  
RP FUNCTION.  
RX MEDLINE-91035239; PubMed-2228961;  
RA Amemura M., Makino K., Shinagawa H., Nakata A.;  
RT "Cross talk to the phosphate regulation of Escherichia coli by Phom  
RT protein: Phom is a histidine protein kinase and catalyzes

RT phosphorylation of Phob and Phom-open reading frame 2.";  
RL J. Bacteriol. 172:6300-6307(1990).  
CC -I- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM CREC/CREB  
CC INVOLVED IN CATABOLIC REGULATION. CREC MAY FUNCTION AS A MEMBRANE-  
CC ASSOCIATED PROTEIN KINASE THAT PHOSPHORYLATES CREB IN RESPONSE TO  
CC ENVIRONMENTAL SIGNALS. CREC CAN ALSO PHOSPHORYLATES PHOB.  
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (PROBABLE).  
CC -I- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE  
CC KINASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M13608; AAA24375.1; -  
DR EMBL; U14003; AAA97295.1; -  
DR EMBL; AE000510; AAC77352.1; -  
DR EMBL; Y00538; CAA68601.1; -  
DR PIR; C25038; RGECEM.  
DR Ecocore; EG10730; crec.  
DR InterPro; IPR000410; -  
DR InterPro; IPR000658; -  
DR Pfam; PF00672; DUP5; 1.  
DR Pfam; PF00512; signal; 1.  
KW Sensory transduction; Transferrase; Kinase; Phosphorylation;  
KW Transmembrane; Inner membrane.  
FT TRANSMEM 7 27  
FT TRANSMEM 147 167 POTENTIAL.  
FT TRANSMEM 184 204 POTENTIAL.  
FT DOMAIN 242 474 TRANSMITTER DOMAIN (POTENTIAL).  
FT MOD\_RES 265 265 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT CONFLICT 77 77 P -> R (IN REF. 2).  
SQ SEQUENCE 474 AA; 52117 MW; E0296C1A56791B95 CRC64;

Query Match 76.3%; Score 29; DB 1; Length 474;  
Best Local Similarity 85.7%; Pred. No. 29;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVNLAEE 7  
|||  
DB 358 EVNVAEE 364

RESULT 6  
DLDH\_CANFA STANDARD: PRT; 509 AA.  
ID DLDH\_CANFA  
AC P49819;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE DIHYDROLIPOAMIDE DEHYDROGENASE, MITOCHONDRIAL PRECURSOR (EC 1.8.1.4).  
GN DLD.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MONGREL; TISSUE=Skeletal muscle;  
RX MEDLINE-95394366; PubMed-7665089;  
RA Martins A.S., Greene L.J., Yoho L.L., Milsted A.;  
RT "The cDNA encoding canine dihydrolipoamide dehydrogenase contains  
RT multiple termination signals.";  
RL Gene 161:253-257(1995).  
RN [2]  
RP SEQUENCE OF 36-49.  
RC TISSUE=Heart;

```

RX MEDLINE-98163340; PubMed-9504812;
RX Dunn M.J., Corbett J.M., Wheeler C.H.;
HSC-2PAGE and the two-dimensional gel electrophoresis database of
RT dog heart proteins."
RL Electrophoresis 18:2795-2802(1997).
CC -1- FUNCTION: LIPOAMIDE DEHYDROGENASE IS A COMPONENT OF THE GLYCINE
CC CLEAVAGE SYSTEM AS WELL AS OF THE ALPHA-KETOACID DEHYDROGENASE
CC COMPLEXES.
CC -1- CATALYTIC ACTIVITY: DIHYDROLIPOAMIDE + NAD(+) = LIPOAMIDE + NADH.
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC OXIDOREDUCTASES CLASS-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U19872; AAA87174.1; -
DR HSCP: P14218; ILPF.
DR HSC-2PAGE; P49819; DGC.
DR InterPro: IPR001100; -
DR InterPro: IPR001327; -
DR Pfam: PF00070; PYR_redox; 1.
DR PRINTS: PR00368; FADPNR.
DR PRINTS: PR00411; PNDRTASE1.
DR PROSITE: PS00076; PYRIDINE_REDOX_1; 1.
DR Redox-active center; Oxidoreductase; NAD; Flavoprotein; FAD;
KM Mitochondrion; transit peptide.
KT TRANSIT 1 MITOCHONDRION.
FT CHAIN 36 509 DIHYDROLIPOAMIDE DEHYDROGENASE.
FT NP_BIND 43 73 FAD (ADP PART) (PROBABLE).
FT DISULFID 80 85 REDOX-ACTIVE.
FT NP_BIND 345 355 FAD (FLAVIN PART) (BY SIMILARITY).
FT ACT_SITE 487 487 BY SIMILARITY.
FT SEQUENCE 509 AA; 54153 MW; BC2553C52AB30E20 CRC64;
SQ

```

```

Query Match 76.3%; Score 29; DB 1; Length 509;
Best Local Similarity 75.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 EYVLAEEF 8
DB 496 EAMLAASF 503

```

```

RESULT 7
DIDH_HUMAN STANDARD; PRT; 509 AA.
ID P09622; Q14167; Q14131;
AC P09622; Q14167; Q14131;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DIHYDROLIPOAMIDE DEHYDROGENASE, MITOCHONDRIAL PRECURSOR (EC 1.6.1.4).
GN DID OR LAD OR PHE3 OR GCSL.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-88087005; PubMed-3693355;
RA Oculakowski G., Robinson B.H.;
RT "Isolation and sequence determination of cDNA clones for porcine and
RT human lipamide dehydrogenase. Homology to other disulfide
RT oxidoreductases."

```

```

RL J. Biol. Chem. 262:17313-17318(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-8814449; PubMed-3278312;
RA Pons G., Raefsky-Estrin C., Carothers D.J., Pepln R.A., Javed A.A.,
RA Jesse B.W., Ganapathi M.K., Samols D., Patel M.S.;
RT "Cloning and cDNA sequence of the dihydrolipoamide dehydrogenase
RT component human alpha-ketoacid dehydrogenase complexes."
RT Proc. Natl. Acad. Sci. U.S.A. 85:1422-1426(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE-94010932; PubMed-8406489;
RA Feigenbaum A.S., Robinson B.H.;
RT "The structure of the human dihydrolipoamide dehydrogenase gene (DLD)
RT and its upstream elements."
RL Genomics 17:376-381(1993).
RN [4]
RP SEQUENCE OF 1-13 FROM N.A.
RX MEDLINE-9306364; PubMed-1332063;
RA Johanning G.L., Morris J.I., Madhusudan K.T., Samols D., Patel M.S.;
RT "Characterization of the transcriptional regulatory region of the
RT human dihydrolipoamide dehydrogenase gene."
RL Proc. Natl. Acad. Sci. U.S.A. 89:10964-10968(1992).
RN [5]
RP VARIANTS GLU-72 AND LEU-488.
RX MEDLINE-93281722; PubMed-8506365;
RA Liu T.-C., Kim H., Arizmendi C., Kitano A., Patel M.S.;
RT "Identification of two missense mutations in a dihydrolipoamide
RT dehydrogenase-deficient patient."
RL Proc. Natl. Acad. Sci. U.S.A. 90:5186-5190(1993).
CC -1- FUNCTION: LIPOAMIDE DEHYDROGENASE IS A COMPONENT OF THE GLYCINE
CC CLEAVAGE SYSTEM AS WELL AS OF THE ALPHA-KETOACID DEHYDROGENASE
CC COMPLEXES.
CC -1- CATALYTIC ACTIVITY: DIHYDROLIPOAMIDE + NAD(+) = LIPOAMIDE + NADH.
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -1- DISEASE: DEFECTS IN E3 RESULTS IN CONGENITAL INFANTILE LACTIC
CC ACIDOSIS.
CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC OXIDOREDUCTASES CLASS-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J03490; AAA59527.1; -
DR EMBL: J03620; AAA5764.1; -
DR EMBL: L13761; AAB01381.1; JOINED.
DR EMBL: L13749; AAB01381.1; JOINED.
DR EMBL: L13750; AAB01381.1; JOINED.
DR EMBL: L13751; AAB01381.1; JOINED.
DR EMBL: L13752; AAB01381.1; JOINED.
DR EMBL: L13753; AAB01381.1; JOINED.
DR EMBL: L13748; AAB01381.1; JOINED.
DR EMBL: L13755; AAB01381.1; JOINED.
DR EMBL: L13759; AAB01381.1; JOINED.
DR EMBL: L13760; AAB01381.1; JOINED.
DR EMBL: L13756; AAB01381.1; JOINED.
DR EMBL: L13757; AAB01381.1; JOINED.
DR EMBL: L13758; AAB01381.1; JOINED.
DR EMBL: M99384; AAA35759.1; -
DR PIR: B28448; DEHULP.
DR HSCP: P14218; ILPF.
DR HSC-2PAGE; P09622; HUMAN.
MTM: 246900; -

```

```

DR MIM: 238331;
DR InterPro: IPR001100;
DR InterPro: IPR001327;
DR Pfam: PF00070; pyr_redox: 1.
DR PRINTS: PR00368; FADPNR.
DR PRINTS: PR00411; PNDRTASE1.
DR PROSITE: PS00076; PYRIDINE_REDOX_1; 1.
DR Redox-active center; Oxidoreductase; NAD; Flavoprotein; FAD;
KM Mitochondrion; Transf. peptide.
FT TRANSIT 1 35 MITOCHONDRION.
FT CHAIN 36 509 DIHYDROLIPOAMIDE DEHYDROGENASE.
FT NP_BIND 43 73 FAD (ADP PART) (PROBABLE).
FT DISULFID 80 85 REDOX-ACTIVE.
FT NP_BIND 345 355 FAD (FLAVIN PART) (BY SIMILARITY).
FT ACT_SITE 487 487 BY SIMILARITY.
FT VARIANT 72 72 K->E (IN DLD DEFICIENCY).
FT VARIANT 488 488 P->L (IN DLD DEFICIENCY).
FT VARIANT 488 488 P->L (IN DLD DEFICIENCY).
FT CONFLICT 104 104 T->K (IN REF. 2).
FT CONFLICT 154 154 G->R (IN REF. 2).
FT CONFLICT 493 493 A->AEA (IN REF. 3).
SQ SEQUENCE 509 AA: 54150 MW: BD1006B52F4369B CRC64;

```

```

Query Match          76.3%; Score 29; DB 1; Length 509;
Best Local Similarity 75.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 EVNLAEF 8
DB 496 EANLAASF 503

```

```

RESULT 8
DLDH_PIG 8 STANDARD: PRT: 509 AA.
AC P09623;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last annotation update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DIHYDROLIPOAMIDE DEHYDROGENASE, MITOCHONDRIAL PRECURSOR (EC 1.8.1.4).
GN DLD OR LAD.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=96823;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88087005; PubMed=3693355;
RA Otulakowski G., Robinson B.H.;
RT "Isolation and sequence determination of cDNA clones for porcine and
RT human lipamide dehydrogenase. Homology to other disulfide
RT oxidoreductases."
RL J. Biol. Chem. 262:17313-17318(1987).
CC -1- FUNCTION: LIPOLIPAMIDE DEHYDROGENASE IS A COMPONENT OF THE GLYCINE
CC CLEAVAGE SYSTEM AS WELL AS OF THE ALPHA-KETOACID DEHYDROGENASE
CC COMPLEXES.
CC -1- CATALYTIC ACTIVITY: DIHYDROLIPOAMIDE + NAD(+) -> LIPOLIPAMIDE + NADH.
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC OXIDOREDUCTASES CLASS-I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: J03489; AAA1069.1;
DR PIR: A28448; DEPLP.
DR HSSP: P14218; ILPE.
DR InterPro: IPR001100;
DR InterPro: IPR001327;
DR Pfam: PF00070; pyr_redox: 1.
DR PRINTS: PR00368; FADPNR.
DR PRINTS: PR00411; PNDRTASE1.
DR PROSITE: PS00076; PYRIDINE_REDOX_1; 1.
DR Redox-active center; Oxidoreductase; NAD; Flavoprotein; FAD;
KM Mitochondrion; Transf. peptide.
FT TRANSIT 1 35 MITOCHONDRION.
FT CHAIN 36 509 DIHYDROLIPOAMIDE DEHYDROGENASE.
FT NP_BIND 43 73 FAD (ADP PART) (PROBABLE).
FT DISULFID 80 85 REDOX-ACTIVE.
FT NP_BIND 345 355 FAD (FLAVIN PART) (BY SIMILARITY).
FT ACT_SITE 487 487 BY SIMILARITY.
SQ SEQUENCE 509 AA: 54185 MW: 38A0469FED071300 CRC64;

```

```

Query Match          76.3%; Score 29; DB 1; Length 509;
Best Local Similarity 75.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 EVNLAEF 8
DB 496 EANLAASF 503

```

```

RESULT 9
GLMS_STRCO 9 STANDARD: PRT: 614 AA.
AC O86781;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GLUCOSAMINE-6-FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE (ISOMERIZING)
DE (EC 2.6.1.16) (HEXOSEPHOSPHATE AMINOTRANSFERASE) (D-FRUCTOSE-6-
DE PHOSPHATE AMINOTRANSFERASE) (GFAT) (L-GLUTAMINE-D-FRUCTOSE-6-PHOSPHATE
DE AMINOTRANSFERASE) (GLUCOSAMINE-6-PHOSPHATE SYNTHASE).
GN GLMS OR SC664.18.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=A3(2);
RA Saunders D., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
RA Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE FIRST STEP IN HEXOSAMINE METABOLISM,
CC CONVERTING FRUCTOSE-6P INTO GLUCOSAMINE-6P USING GLUTAMINE AS A
CC NITROGEN SOURCE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-GLUTAMINE + D-FRUCTOSE 6-PHOSPHATE ->
CC L-GLUTAMATE + D-GLUCOSAMINE 6-PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-2
CC GATASE DOMAIN.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.
CC GLMS SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL031317; CAA20396.1;
DR InterPro: IPR000583;
DR InterPro: IPR001347;

```

DR Pfam: PF00310: GATase\_2; 1.  
 DR Pfam: PF01380: SIS; 2.  
 DR PROSITE: PS00443: GATASE\_TYPE\_II; 1.  
 KN Trasnferase: Amino transferase: Glutamine amidotransferase.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 609 609 ISOMERISATION FRU-6P (BY SIMILARITY).  
 FT DOMAIN 1 183 GLUTAMINE AMIDOTRANSFERASE.  
 FT SEQUENCE 614 AA; 65429 MW; 13P3B099E457262A CRC64;  
 SQ

Query Match 76.3%; Score 29; DB 1; Length 614;  
 Best Local Similarity 75.0%; Pred. No. 38;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 DB 325 EVELASEF 332

RESULT 10  
 GLMS\_MYCTU STANDARD; PRT; 623 AA.  
 ID GLMS\_MYCTU  
 AC 006253; O33274;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE [ISOMERIZING]  
 DE (EC 2.6.1.16) (HEXOSEPHOSPHATE AMINOTRANSFERASE) (D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE) (GFRAT) (L-GLUTAMINE-D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE) (GLUCOSAMINE-6-PHOSPHATE SYNTHASE).  
 GN GLMS OR RV3436C OR MTCY77.08C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Mycobacterium.  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID-1773;  
 RX [1]  
 RA SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RA Schaeffer M.L., Bessa G.S., Belisle J.T., Inamine J.M.;  
 RT "Biochemical and genetic definition of effects of mannose on mycobacteria."  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-H37RV;  
 RX MEDLINE-98295987; PubMed-9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RA Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."  
 RT Nature 393:537-544 (1998).  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RA McKendree W.L., Schuster S.M., Richards N.;  
 RT "Structure of a recombinant glms protein from Mycobacterium tuberculosis expressed in E. coli."  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CATALYZES THE FIRST STEP IN HEXOSAMINE METABOLISM, CONVERTING FRUCTOSE-6P INTO GLUCOSAMINE-6P USING GLUTAMINE AS A NITROGEN SOURCE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: L-GLUTAMINE + D-FRUCTOSE 6-PHOSPHATE = L-GUTAMATE + D-GLUCOSAMINE 6-PHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-2 GATASE DOMAIN.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.

CC GLMS SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF002814; AAB69988.1; -  
 CC EMBL: Z95389; CAB08685.1; -  
 CC EMBL: AJ000333; CA04007.1; ALT\_INIT.  
 CC Trasnferase: Amino transferase: Glutamine amidotransferase.  
 DR Trasnferase: Amino transferase: Glutamine amidotransferase.  
 DR InterPro: IPR001347; -  
 DR Pfam: PF00310: GATase\_2; 1.  
 DR PROSITE: PS00443: GATASE\_TYPE\_II; 1.  
 KN Trasnferase: Amino transferase: Glutamine amidotransferase.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 618 618 ISOMERISATION FRU-6P (BY SIMILARITY).  
 FT DOMAIN 1 188 GLUTAMINE AMIDOTRANSFERASE.  
 FT SEQUENCE 623 AA; 67440 MW; 32C54B098E2C8360 CRC64;  
 SQ

Query Match 76.3%; Score 29; DB 1; Length 623;  
 Best Local Similarity 75.0%; Pred. No. 39;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 DB 334 EVELASEF 341

RESULT 11  
 GLMS\_MYCLE STANDARD; PRT; 624 AA.  
 ID GLMS\_MYCLE  
 AC P40831;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE [ISOMERIZING]  
 DE (EC 2.6.1.16) (HEXOSEPHOSPHATE AMINOTRANSFERASE) (D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE) (GFRAT) (L-GLUTAMINE-D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE) (GLUCOSAMINE-6-PHOSPHATE SYNTHASE).  
 GN GLMS OR B229\_C3\_238.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Mycobacterium.  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID-1769;  
 RX [1]  
 RA SEQUENCE FROM N.A.  
 RP Smith D.R., Robison K.;  
 RT Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CATALYZES THE FIRST STEP IN HEXOSAMINE METABOLISM, CONVERTING FRUCTOSE-6P INTO GLUCOSAMINE-6P USING GLUTAMINE AS A NITROGEN SOURCE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: L-GLUTAMINE + D-FRUCTOSE 6-PHOSPHATE = L-GUTAMATE + D-GLUCOSAMINE 6-PHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-2 GATASE DOMAIN.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.  
 CC GLMS SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>

CC or send an email to [license@lsb.sib.ch](mailto:license@lsb.sib.ch).  
 CC -----  
 DR EMBL: U00020: AAA17307.1; -  
 DR HSSP: P17169: IMOR.  
 DR MEROPS: CA4.971; -  
 DR InterPro: IPR000583; -  
 DR InterPro: IPR001347; -  
 DR Pfam: PF00310: GATase\_2; 1.  
 DR Pfam: PF01380: SIS; 2.  
 DR PROSITE: PS00443: GATASE TYPE II: FALSE NEG.  
 DR Transferrase: Aminoctransferase, Glutamine amidotransferase.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 619 619 ISOMERISATION FRU-6P (BY SIMILARITY).  
 FT DOMAIN 1 192 GLUTAMINE AMIDOTRANSFERASE.  
 SQ SEQUENCE 624 AA: 67371 MW: 118967AEAD7E76D8 CRC64;

Query Match  
 Best Local Similarity 76.3%; Score 29; DB 1; Length 624;  
 Best Local Similarity 75.0%; Pred. No. 39;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 DB 335 EVELASEF 342

RESULT 12  
 GLMS\_MYCSM  
 ID GLMS\_MYCSM STANDARD: PRT: 627 AA.  
 AC 068956;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GLUCOSAMINE-6-PHOSPHATE AMINOTRANSFERASE [ISOMERIZING]  
 DE (EC 2.6.1.16) (HEXOSEPHOSPHATE AMINOTRANSFERASE) (D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE) (G6P) (L-GLUTAMINE-D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE) (GLUCOSAMINE-6-PHOSPHATE SYNTHASE).  
 GN GLMS.  
 OS Mycobacterium smegmatis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1772;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 700084 / MC(2)155;  
 RA Scheffer M.L., Besta G.S., Belisle J.T., Inamine J.M.;  
 RT "Biochemical and genetic definition of effects of mannoseamine on mycobacteria";  
 RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CATALYZES THE FIRST STEP IN HEXOSAMINE METABOLISM, CONVERTING FRUCTOSE-6P INTO GLUCOSAMINE-6P USING GLUTAMINE AS A NITROGEN SOURCE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: L-GLUTAMINE + D-FRUCTOSE 6-PHOSPHATE -  
 CC L-GALUTAMATE + D-GLUCOSAMINE 6-PHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-2 GATASE DOMAIN.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.  
 CC GLMS SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----  
 DR EMBL: AF058788: AAC14295.1; -  
 DR InterPro: IPR000583; -  
 DR InterPro: IPR001347; -  
 DR Pfam: PF00310: GATase\_2; 1.

DR Pfam: PF01380: SIS; 2.  
 DR PROSITE: PS00443: GATASE TYPE II: 1.  
 DR Transferrase: Aminoctransferase, Glutamine amidotransferase.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 622 622 ISOMERISATION FRU-6P (BY SIMILARITY).  
 FT DOMAIN 1 192 GLUTAMINE AMIDOTRANSFERASE.  
 SQ SEQUENCE 627 AA: 67771 MW: AA82F590871F926E CRC64;

Query Match  
 Best Local Similarity 76.3%; Score 29; DB 1; Length 627;  
 Best Local Similarity 75.0%; Pred. No. 39;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 DB 338 EVELASEF 345

RESULT 13  
 VARF\_BPP22  
 ID VARF\_BPP22 STANDARD: PRT: 47 AA.  
 AC P14112;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE ACCESSORY RECOMBINATION FUNCTION PROTEIN.  
 GN ARF.  
 OS Bacteriophage P22.  
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.  
 OX NCBI\_TaxID=10754;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89293845; PubMed=2738922;  
 RA Semerjian A.V., Malloy D.C., Poteete A.R.;  
 RT "Genetic structure of the bacteriophage P22 PL operon";  
 RL J. Mol. Biol. 207:1-13(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kropinski A.M.B., Vanderbyl C.S.;  
 RT "The completed sequence of genome of Salmonella phage P22.";  
 RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ARF MAY BE RECOMBINATION-RELATED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----  
 DR EMBL: X15637: CAA33652.1; -  
 DR EMBL: AF217253: AAF75015.1; -  
 DR PIR: S04249: W2BP22.  
 SQ SEQUENCE 47 AA: 5517 MW: 2C8682C7B4610FE3 CRC64;

Query Match  
 Best Local Similarity 73.7%; Score 28; DB 1; Length 47;  
 Best Local Similarity 71.4%; Pred. No. 4;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVNLAAEF 7  
 DB 25 QINLAAEF 31

RESULT 14  
 YRB2\_HALCU  
 ID YRB2\_HALCU STANDARD: PRT: 68 AA.  
 AC P17104;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)

01-AUG-1990 (rel. 15, Last annotation update)  
 DE HYPOTHEORETICAL 23 KDA PROTEIN IN RIBOSOMAL PROTEIN GENE CLUSTER 5' REGION  
 DE (PROTEIN NAB).  
 OS Halobacterium cutirubrum.  
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium.  
 OX NCBI\_TaxID=2240;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=34001;  
 RX MEDLINE=89305527; PubMed=2743981;  
 RA Shlimun L.C., Dennis P.P.;  
 RT "Characterization of the L11, L1, L10 and L12 equivalent ribosomal  
 RT protein gene cluster of the halophilic archaeobacterium Halobacterium  
 RT cutirubrum.";  
 RL EMBL J. 8:1225-1235(1989).  
 CC -1- SIMILARITY: SOME, TO RESTRICTION ENDONUCLEASES ECO RI AND PST I.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X15078; CAA33177.1; -  
 CC PIR: S04117; QOH5NB.  
 DR Hypothetical Protein.  
 KM SEQUENCE 68 AA; 7539 MW; CC68625E29D3A91F CRC64;  
 SO

Query Match 73.7%; Score 28; DB 1; Length 68;  
 Best Local Similarity 75.0%; Pred. No. 6;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYNLAEE 8  
 :|||  
 DB 24 EVDLEAEF 31

RESULT 15  
 ROC\_XENLA STANDARD; PRT; 282 AA.  
 ID ROC\_XENLA  
 AC P19600;  
 DT 01-FEB-1991 (rel. 17, Created)  
 DT 01-FEB-1991 (rel. 17, Last sequence update)  
 DT 01-OCT-2000 (rel. 40, Last annotation update)  
 DE HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN C (HNRP CORE PROTEIN C).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89071757; PubMed=2904678;  
 RA Preugschat F, Wold B;  
 RT "Isolation and characterization of a Xenopus laevis C protein cDNA:  
 RT structure and expression of a heterogeneous nuclear ribonucleoprotein  
 RT core protein.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9669-9673(1988).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN NUCLEOSOME ASSEMBLY BY NEUTRALIZING  
 CC BASIC PROTEINS SUCH AS A AND B CORE HNRPNS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; COMPONENT OF RIBONUCLEOSOMES.  
 CC -1- PTM: PHOSPHORYLATED (PROBABLE).  
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: J03831; AAA60937.1; -  
 CC PIR: A31765; A31765.  
 DR HSSP: P09012; 1AUD.  
 DR InterPro: IPR000504; -  
 DR Pfam: PF00076; Rrm; 1.  
 DR PROSITE: PS50102; RRM; 1.  
 DR PROSITE: PS00030; RRM; RNP\_1; 1.  
 KM Nuclear protein; RNA-binding; Ribonucleoprotein; Phosphorylation.  
 FT DOMAIN 17 88  
 FT DOMAIN 147 282  
 FT DOMAIN 178 240  
 FT MOD\_RES 240 240  
 FT MOD\_RES 256 256  
 FT MOD\_RES 267 267  
 SQ SEQUENCE 282 AA; 30950 MW; 7373FA46F8C85413 CRC64;

Query Match 73.7%; Score 28; DB 1; Length 282;  
 Best Local Similarity 71.4%; Pred. No. 28;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYNLAEE 7  
 :|||  
 DB 82 DINTLAE 88

Search completed: September 6, 2001, 16:51:11  
 Job time: 815 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:54 ; Search time 134.15 Seconds  
(without alignments)  
4.543 Million cell updates/sec

Title: US-09-603-713-28

Perfect score: 38

Sequence: 1 EVNLAAEF 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: PIR\_68:\*

1: p1r1:.\*  
2: p1r2:.\*  
3: p1r3:.\*  
4: p1r4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	84.2	538	2 E96492	hypothetical prote
2	31	81.6	466	2 F64242	hypothetical prote
3	31	81.6	466	2 F84632	hypothetical prote
4	30	78.9	280	2 D84358	phosphonates trans
5	30	78.9	283	2 T01454	hypothetical prote
6	30	78.9	295	2 D86255	hypothetical prote
7	30	78.9	2493	2 H81324	probable MCP-domai
8	29	76.3	282	2 S45734	probable membrane
9	29	76.3	328	2 T03227	hypothetical prote
10	29	76.3	387	2 C70368	conserved hypotet
11	29	76.3	387	2 B55164	scnl protein - fls
12	29	76.3	416	2 S76310	hypothetical prote
13	29	76.3	430	2 S77393	cell division prote
14	29	76.3	474	1 RGECEM	sensor protein cre
15	29	76.3	474	2 G86139	hypothetical prote
16	29	76.3	509	1 DEHULP	dihydrolipoamide d
17	29	76.3	509	1 DEHULP	dihydrolipoamide d
18	29	76.3	509	1 JC4241	dihydrolipoamide d
19	29	76.3	529	2 T42584	legument protein d
20	29	76.3	615	2 T35569	glutamine--fructos
21	29	76.3	624	2 B70976	glutamine--fructos
22	29	76.3	625	2 S72993	glutamine--fructos
23	28	73.7	47	1 W2BP22	art protein - phag
24	28	73.7	68	1 OQHSNB	hypothetical prote
25	28	73.7	68	2 B84267	transcriptional regu
26	28	73.7	237	2 B69785	transcriptional regu
27	28	73.7	253	2 S77778	probable NH(3)-dep
28	28	73.7	282	2 A31765	heterogeneous ribo
29	28	73.7	289	2 S12619	Na+-transporting A

30	28	73.7	290	2 A26885	heterogeneous nucl
31	28	73.7	303	2 C34504	heterogeneous ribo
32	28	73.7	315	2 C69763	ferichrome ABC tr
33	28	73.7	328	2 T15061	hypothetical prote
34	28	73.7	463	2 J50376	hypothetical 53.0K
35	28	73.7	600	2 D83683	L-glutamine-D-fruc
36	28	73.7	676	2 B72071	ribonuclease fam11
37	28	73.7	676	2 D86553	ribonuclease fam11
38	28	73.7	734	2 C69691	GTP pyrophosphokin
39	28	73.7	1131	2 S52266	FUN30 protein - ye
40	27	71.1	161	1 G64814	molybdenum cofacto
41	27	71.1	161	1 F85386	molybdopterin bios
42	27	71.1	215	2 A29318	ubiquinol--cytochr
43	27	71.1	216	2 B85870	probable phosphata
44	27	71.1	222	2 C65001	hypothetical prote
45	27	71.1	228	2 S76876	hypothetical prote

#### ALIGNMENTS

RESULT 1  
E96492  
hypothetical protein F5A13.3 [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: E96492  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart,  
ansen, N.F.; Hughes, B.; Hutzler, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marzla  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: E96492  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-538 <STO>  
A:Cross-references: GB:A8005173; NID:99802762; PIDN:AAF99831.1; GSPDB:GM00141  
C:Genetics:  
A:Gene: F5A13.3  
A:Map position: 1

Query Match 84.2%; Score 32; DB 2; Length 538;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVNLAAE 7  
DB 532 EVNLAAE 538

RESULT 2  
F64242  
hypothetical protein MG385 - Mycoplasma genitalium  
C:Species: Mycoplasma genitalium  
C:Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 21-Jan-2000  
C:Accession: F64242  
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.  
M.; Fuhmann, J.; Nguyen, D.; Uterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,  
C.A.; Venter, J.C.  
Science 270, 397-403, 1995  
A:Title: The minimal gene complement of Mycoplasma genitalium.  
A:Reference number: A64200; MUID:96026346  
A:Accession: F64242  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-236 <TRIG>

A:Cross-references: GB:U93723; GB:L43967; NID:g1046092; PID:g1046096; TIGR:M385  
 A:Experimental source: strain G-37  
 C:Genetics:

A:Genetic code: SGC3  
 C:Superfamily: glycerophosphodiester phosphodiesterase

Query Match 81.6%; Score 31; DB 2; Length 236;  
 Best Local Similarity 75.0%; Pred. No. 10;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVNLAEEF 8  
 Db 74 KVNLAEEF 81

RESULT 3  
 F84632  
 hypothetical protein At2g24100 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: F84632  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.;  
 euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: F84632

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-466 <STO>

A:Cross-references: GB:AB002093; NID:g4115371; PIDN:AD003372.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g24100

A:Map position: 2

Query Match 81.6%; Score 31; DB 2; Length 466;  
 Best Local Similarity 75.0%; Pred. No. 22;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAEEF 8  
 Db 252 EYNLAEEF 259

RESULT 4  
 DB4358  
 phosphonates transport ATP-binding [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: DB4358  
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.  
 ; Letlhauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

Jung, K.H.; Alam, M.; Freitas, T.

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li

A:Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; MUID:20504483

A:Accession: DB4358

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-280 <STO>

A:Cross-references: GB:AE004437; NID:g10561501; PIDN:AG20232.1; GSPDB:GN00138

C:Genetics:

A:Gene: phnc

Query Match 78.9%; Score 30; DB 2; Length 280;  
 Best Local Similarity 75.0%; Pred. No. 22;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAEEF 8  
 Db 198 QVNLAEEF 205

RESULT 5  
 T01454  
 hypothetical protein F2401.16 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 23-Mar-2001

C:Accession: T01454

R:Shim, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.;

eologis, A.; Ecker, J.R.

submitted to the EMBL Data Library, January 1998

A:Description: Genomic sequence for Arabidopsis thaliana BAC F2401.

A:Reference number: 214211

A:Accession: T01454

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-283 <SHI>

A:Cross-references: EMBL:AC003113; NID:g2689438; PID:g2781360; GSPDB:GN00059; ATSP:F2

C:Genetics:

A:Gene: ATSP:F2401.16

A:Map position: 1

A:introns: 124/1

Query Match 78.9%; Score 30; DB 2; Length 283;  
 Best Local Similarity 75.0%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVNLAEEF 8  
 Db 159 EYNLAEEF 166

RESULT 6  
 DB6255  
 hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: DB6255

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: DB6255

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-295 <STO>

A:Cross-references: GB:AB005172; NID:g3157929; PIDN:AC17612.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 78.9%; Score 30; DB 2; Length 295;  
 Best Local Similarity 75.0%; Pred. No. 23;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EVNLAEEF 8  
 Db 159 EYNLAEEF 166

RESULT 7  
 H81324

probable MCP-domain signal transduction protein Cj1190c [imported] - Campylobacter jejuni  
 C:Species: Campylobacter jejuni  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 31-Mar-2000  
 C:Accession: H81324  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chilling  
 C.W.; Ouali, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barré  
 Nature 403, 665-668, 2000  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyf  
 A:Reference number: A81250; MUID:20150912  
 A:Accession: H81324  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-459 <PAR>  
 A:Cross-references: GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CAB73444.1; PID:g696862  
 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: Cj1190c

Query Match 78.9%; Score 30; DB 2; Length 459;

Best Local Similarity 75.0%; Pred. No. 38;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 EVNLAEF 8  
 :|||:  
 Db 407 EMLAKEF 414

RESULT 8  
 S45734  
 probable membrane protein YBL004w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBL0101

C:Species: Saccharomyces cerevisiae

C:Date: 26-Aug-1994 #sequence\_revision 09-Sep-1994 #text\_change 07-May-1999

C:Accession: S45734; S45733; S44556; S37318

R:Rieger, M.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45734

A:Accession: S45734

A:Molecule type: DNA

A:Residues: 1-2066 <RIE>

A:Cross-references: EMBL:Z35765; MIPS:YBL004w

A:Experimental source: strain S288C

R:Lohan, A.J.E.; Wolfe, K.H.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45730

A:Accession: S45733

A:Molecule type: DNA

A:Residues: 1214-2493 <LOH>

A:Cross-references: EMBL:Z35765; MIPS:YBL004w

A:Experimental source: strain S288C

R:Wolfe, K.H.; Lohan, A.J.E.

Yeast 10(Suppl.A), S41-S46, 1994

A:Title: Sequence around the centromere of Saccharomyces cerevisiae chromosome II: sim1

A:Reference number: S44556; MUID:94378721

A:Accession: S44556

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1214-2493 <WOL>

A:Cross-references: EMBL:Z26494; NID:g403311; PID:g403312

A:Experimental source: strain S288C

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1993

C:Genetics:

A:Map position: 2L

C:Keywords: transmembrane protein

F:165-182/Domain: Transmembrane #status predicted <TM1>

F:150-366/Domain: Transmembrane #status predicted <TM2>

F:656-672/Domain: Transmembrane #status predicted <TM3>

F:1021-1037/Domain: Transmembrane #status predicted <TM4>

F:1170-1187/Domain: Transmembrane #status predicted <TM5>

F:1195-1211/Domain: Transmembrane #status predicted <TM6>

F:1606-1632/Domain: Transmembrane #status predicted <TM7>

F:1656-1674/Domain: Transmembrane #status predicted <TM8>

F:1763-1779/Domain: transmembrane #status predicted <TM9>  
 F:1910-1926/Domain: transmembrane #status predicted <TM10>  
 F:2093-2109/Domain: transmembrane #status predicted <TM11>  
 F:2355-2371/Domain: transmembrane #status predicted <TM12>

Query Match 78.9%; Score 30; DB 2; Length 2493;

Best Local Similarity 62.5%; Pred. No. 2,6e+02;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EVNLAEF 8  
 :|||:  
 Db 56 EMLSAKF 63

RESULT 9

T03227

hypothetical protein - Streptomyces hygroscopicus

C:Species: Streptomyces hygroscopicus

C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 22-Oct-1999

C:Accession: T03227

R:Ruan, X.; Stassi, D.; Lax, S.; Katz, L.

Gene 203, 1-9, 1997

A:Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus AT

A:Reference number: Z14848; MUID:98085969

A:Accession: T03227

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-282 <RNA>

A:Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38067.1; PID:g2624954

A:Experimental source: ATCC 29253

Query Match 76.3%; Score 29; DB 2; Length 282;

Best Local Similarity 85.7%; Pred. No. 39;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EVNLAEF 7  
 :|||:  
 Db 111 QVNLAEE 117

RESULT 10

C70368

conserved hypothetical protein aq\_778 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 20-Sep-1999

C:Accession: C70368

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666

A:Accession: C70368

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-328 <AOF>

A:Cross-references: GB:AE000707; NID:g2983342; PIDN:AAC06942.1; PID:g2983354; GB:AE00

A:Experimental source: strain VFS

C:Genetics:

A:Gene: aq\_778

C:superfamily: Escherichia coli ybhk protein

Query Match 76.3%; Score 29; DB 2; Length 328;

Best Local Similarity 85.7%; Pred. No. 46;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 VNLAEEF 8  
 :|||:  
 Db 133 VNLAEEF 139

RESULT 11  
 scnl protein - fission yeast (Schizosaccharomyces pombe)  
 A:Residues: 1-430 <KAN>  
 A:Cross-references: EMBL:D90906; GB:AB001339; NID:q1652492; PIDN:BA17496.1; PID:q165  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 23-Mar-1995 #sequence\_revision 23-Mar-1995 #text\_change 29-Oct-1999  
 C:Accession: B55164  
 R:Samelima, I.; Yanagida, M.  
 J. Cell Biol. 127, 1655-1670, 1994  
 A:Title: Bypassing anaphase by fission yeast cut9 mutation: requirement of cut9(+) to in  
 A:Reference number: A55164; MUID:95096177  
 A:Accession: B55164  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-387 <SAM>  
 A:Cross-references: GB:D31845; NID:9633128; PIDN:BA06631.1; PID:d1007203; PID:q633129  
 C:Genetics:  
 A:Gene: scnl

Query Match	76.3%	Score 29;	DB 2;	Length 387;
Best Local Similarity	75.0%	Pred. No. 55;		
Matches	6;	Conservative 1;	Mismatches 1;	Indels 0;
			Gaps 0;	

OY 1 EVNLAER 8  
 |||||  
 Db 223 QVRLAER 230

RESULT 12  
 S76310  
 hypothetical protein - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 C:Accession: S76310  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S76310  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-416 <KAN>  
 A:Cross-references: EMBL:D64000; GB:AB001339; NID:q1001484; PIDN:BA10162.1; PID:q100153  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match	76.3%	Score 29;	DB 2;	Length 416;
Best Local Similarity	85.7%	Pred. No. 60;		
Matches	6;	Conservative 1;	Mismatches 0;	Indels 0;
			Gaps 0;	

OY 1 EVNLAER 7  
 |||||  
 Db 245 EVNLAAR 251

RESULT 13  
 S77393  
 cell division protein ftsz - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 C:Accession: S77393  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
 A:Reference number: S74322; MUID:97061201  
 A:Accession: S77393  
 A:Status: nucleotide acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-430 <KAN>  
 A:Cross-references: EMBL:D90906; GB:AB001339; NID:q1652492; PIDN:BA17496.1; PID:q165  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 23-Mar-1995 #sequence\_revision 23-Mar-1995 #text\_change 29-Oct-1999  
 C:Accession: B55164  
 R:Samelima, I.; Yanagida, M.  
 J. Cell Biol. 127, 1655-1670, 1994  
 A:Title: Bypassing anaphase by fission yeast cut9 mutation: requirement of cut9(+) to in  
 A:Reference number: A55164; MUID:95096177  
 A:Accession: B55164  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-387 <SAM>  
 A:Cross-references: GB:D31845; NID:9633128; PIDN:BA06631.1; PID:d1007203; PID:q633129  
 C:Genetics:  
 A:Gene: scnl

Query Match	76.3%	Score 29;	DB 2;	Length 430;
Best Local Similarity	85.7%	Pred. No. 62;		
Matches	6;	Conservative 1;	Mismatches 0;	Indels 0;
			Gaps 0;	

OY 1 EVNLAER 7  
 |||||  
 Db 329 EVNVAAR 335

RESULT 14

RGBCFM  
 sensor protein crec (EC 2.7.3.-) - Escherichia coli

N:Alternate names: pho regulon positive regulatory protein crec  
 C:Species: Escherichia coli  
 C:Date: 31-Mar-1988 #sequence\_revision 17-Oct-1997 #text\_change 16-Jul-1999  
 C:Accession: S56623; F65255; C25038; S03764  
 R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 23, 2105-2119, 1995  
 A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from  
 A:Reference number: S56314; MUID:95334362  
 A:Accession: S56623  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-474 <BUR>  
 A:Cross-references: EMBL:U14003; NID:q1263172; PIDN:AA97295.1; PID:q537239  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
 A.; Rose, D.J.; Mau, B.; Zhao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:9742617  
 A:Accession: F65255  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-474 <BLAT>  
 A:Cross-references: GB:AB000510; GB:U00096; NID:q1790858; PIDN:AACT7352.1; PID:q17908

A:Experimental source: strain K-12, substrain MG1655  
 R:Amemura, M.; Makino, K.; Shinagawa, H.; Nakata, A.  
 J. Bacteriol. 168, 294-302, 1986  
 A:Title: Nucleotide sequence of the phoM region of Escherichia coli: four open readin  
 A:Reference number: A25038; MUID:87008393  
 A:Accession: C25038

A:Molecule type: DNA  
 A:Residues: 1-76; 'P', '78-474 <AME>  
 A:Cross-references: EMBL:M13608; NID:q147248; PIDN:AAA24375.1; PID:q147251  
 R:Drury, L.S.; Buxton, R.S.  
 Mol. Microbiol. 2, 109-119, 1988  
 A:Title: Identification and sequencing of the Escherichia coli cet gene which codes f  
 A:Reference number: S03764; MUID:88216172  
 A:Accession: S03764

A:Molecule type: DNA  
 A:Residues: 381-474 <DRU>  
 A:Cross-references: EMBL:Y00538; NID:q41102; PIDN:CAA68601.1; PID:q809670  
 C:Genetics:  
 A:Gene: crec; phoM  
 A:Map position: 100 min  
 C:Superfamily: enz protein; sensor histidine kinase homology  
 C:Keywords: autophosphorylation; phosphohistidine; phosphoprotein; phosphotransferase  
 F:235-470/Domain: sensor histidine kinase homology <SHK>  
 F:265/Binding site: phosphate (His) (covalent) (by autophosphorylation) #status predi

Query Match 76.3%; Score 29; DB 1; Length 474;  
Best Local Similarity 85.7%; Pred. No. 69;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYNLAEE 7  
|||:||||  
Db 358 EYNVAAE 364

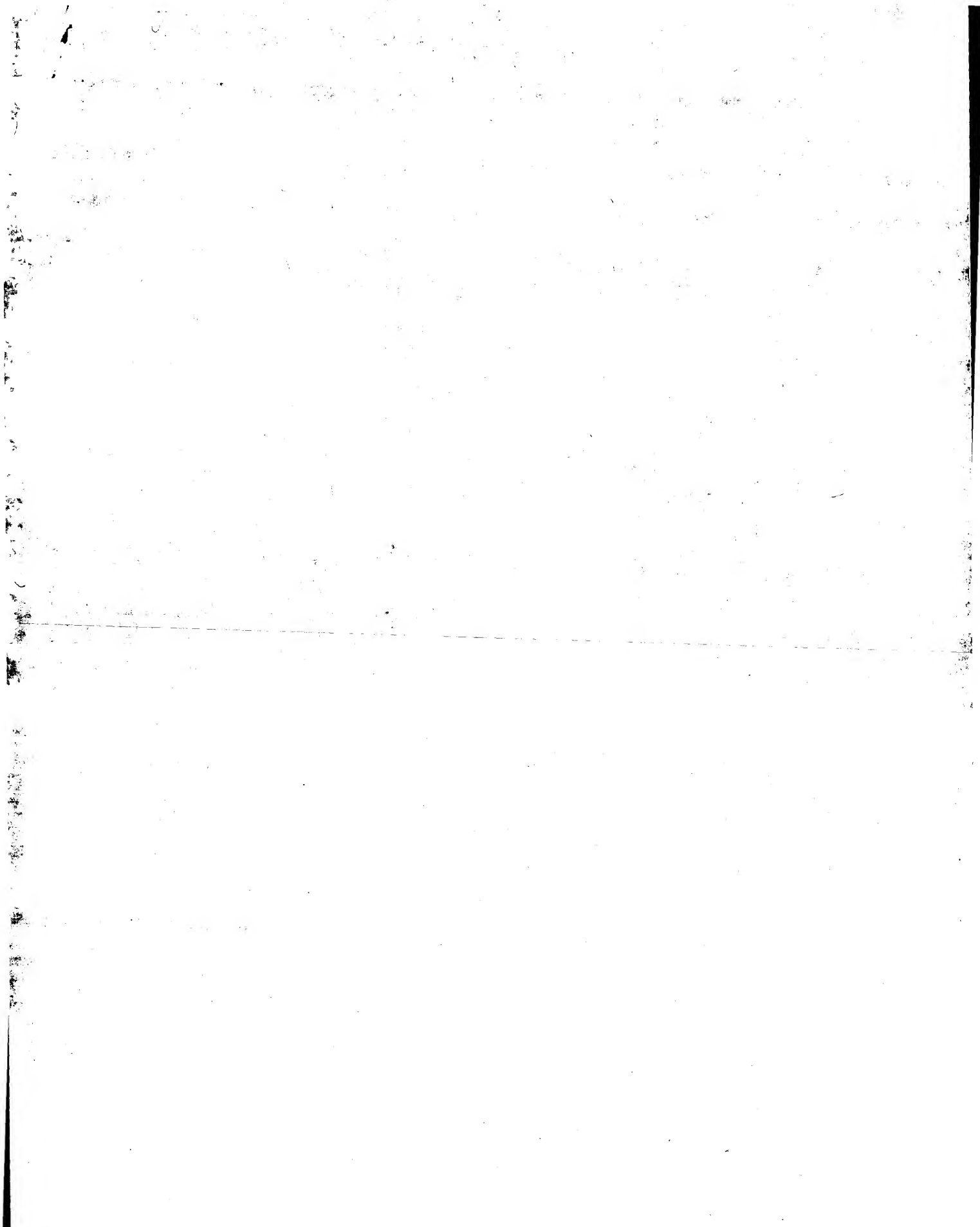
## RESULT 15

G86139  
hypothetical protein crec [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: G86139  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: AB5480; MUID:21074935; PMID:11206551  
A:Accession: G86139  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-474 <STG>  
A:Cross-references: GB:AE005174; NID:q12519431; PIDN:AAG59579.1; GSPDB:GN00145; UWGP:260  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: crec

Query Match 76.3%; Score 29; DB 2; Length 474;  
Best Local Similarity 85.7%; Pred. No. 69;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EYNLAEE 7  
|||:||||  
Db 358 EYNVAAE 364.

Search completed: September 6, 2001, 16:45:55  
Job time: 499 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:39 : Search time 113.12 Seconds  
(without alignments)  
1.456 Million cell updates/sec

Title: US-09-603-713-28

Perfect score: 38

Sequence: 1 EVNLAEF 8

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Issued Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	84.2	9	3	US-08-802-981-219
2	32	84.2	10	2	US-08-659-984A-19
3	32	84.2	10	4	US-08-660-531-19
4	32	84.2	11	5	PCR-US94-07043A-3
5	32	84.2	21	2	US-08-659-984A-18
6	32	84.2	21	3	US-08-802-981-112
7	32	84.2	21	4	US-08-660-531-18
8	32	84.2	30	2	US-08-659-984A-17
9	32	84.2	30	4	US-08-660-531-17
10	32	84.2	33	2	US-08-659-984A-16
11	32	84.2	33	4	US-08-660-531-16
12	32	84.2	42	2	US-08-659-984A-15
13	32	84.2	42	4	US-08-660-531-15
14	32	84.2	103	3	US-08-339-708A-12
15	32	84.2	506	2	US-08-659-984A-21
16	32	84.2	506	4	US-08-660-531-21
17	29	76.3	657	4	US-09-306-593-2
18	28	73.7	386	4	US-09-085-199B-2
19	28	73.7	914	4	US-09-085-199B-4
20	28	73.7	1090	4	US-09-085-199B-4
21	27	71.1	27	1	US-08-141-324-12
22	27	71.1	27	1	US-08-541-902-12
23	27	71.1	47	1	US-08-415-751-12
24	27	71.1	226	4	US-09-299-378-2
25	27	71.1	341	3	US-08-725-459B-44
26	27	71.1	434	3	US-08-725-459B-42
27	27	71.1	437	2	US-08-883-515-4

28	27	71.1	493	3	US-09-090-808-4	Sequence 4, Appl1
29	27	71.1	621	2	US-08-419-652-4	Sequence 4, Appl1
30	27	71.1	660	1	US-08-248-532-3	Sequence 3, Appl1
31	27	71.1	660	2	US-08-419-652-3	Sequence 3, Appl1
32	27	71.1	662	1	US-08-248-532-2	Sequence 2, Appl1
33	27	71.1	662	2	US-08-419-652-2	Sequence 2, Appl1
34	27	71.1	662	2	US-08-685-118-4	Sequence 4, Appl1
35	27	71.1	662	2	US-08-915-495-4	Sequence 4, Appl1
36	27	71.1	662	2	US-08-914-520-4	Sequence 4, Appl1
37	27	71.1	662	3	US-08-789-350-2	Sequence 2, Appl1
38	27	71.1	1456	1	US-08-803-972-2	Sequence 2, Appl1
39	27	71.1	1456	1	US-08-803-972-2	Sequence 2, Appl1
40	26	68.4	9	3	US-08-802-981-220	Sequence 220, App
41	26	68.4	9	3	US-08-802-981-223	Sequence 223, App
42	26	68.4	17	3	US-09-101-146-52	Sequence 52, Appl
43	26	68.4	21	3	US-08-802-981-113	Sequence 113, App
44	26	68.4	21	3	US-08-802-981-116	Sequence 116, App
45	26	68.4	151	3	US-09-142-514-3	Sequence 3, Appl1

## ALIGNMENTS

RESULT 1  
US-08-802-981-219  
: Sequence 219, Application US/08802981  
: Patent No. 6037137  
: GENERAL INFORMATION:  
: APPLICANT: Komoriya, Akira  
: APPLICANT: Packard, Beverly S.  
: TITLE OF INVENTION: Compositions for the Detection of Enzyme  
: TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
: NUMBER OF SEQUENCES: 231  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Townsend and Townsend and Crew LLP  
: STREET: Two Embarcadero Center, Eighth Floor  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94111-3834  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/802,981  
: FILING DATE: 20-FEB-1997  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Hunter, Tom  
: REGISTRATION NUMBER: 38,498  
: REFERENCE/DOCKET NUMBER: 016865-0003000US  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (415) 576-0200  
: TELEFAX: (415) 576-0300  
: INFORMATION FOR SEQ ID NO: 219:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 9 amino acids  
: TYPE: amino acid  
: STRANDEDNESS:  
: TOPOLOGY: linear  
: MOLECULE TYPE: peptide  
: US-08-802-981-219

Query Match 84.2% Score 32: DB 3: Length 9:  
Best Local Similarity 87.5% Pred. No. 1.4e+05:  
Matches 7: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 1 EVNLAEF 8  
||| |||  
DB 2 EVNLAEF 9

RESULT 2  
US-08-659-984A-19  
Sequence 19, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-0028100S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Region  
LOCATION: one-of(1)  
OTHER INFORMATION: /note= "N-terminal Ser is acetylated."  
US-08-659-984A-19

Query Match 84.2%; Score 32; DB 2; Length 10;  
Best Local Similarity 87.5%; Pred. No. 0.12;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYNLAEEF 8  
||| |||  
Db 2 EYNLAEEF 9

RESULT 3  
US-08-660-531-19  
Sequence 19, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-0022100S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Region  
LOCATION: one-of(1)  
OTHER INFORMATION: /note= "N-terminal Ser is acetylated."  
US-08-660-531-19

Query Match 84.2%; Score 32; DB 4; Length 10;  
Best Local Similarity 87.5%; Pred. No. 0.12;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EYNLAEEF 8  
||| |||  
Db 2 EYNLAEEF 9

RESULT 4  
PCT-US94-07043A-3  
Sequence 3, Application PC/TUS9407043A  
GENERAL INFORMATION:  
APPLICANT: Tamburini, Paul P.; Benz, G nter; H bich,  
APPLICANT: Dieter; Dreyer, Robert N.; Koendig, Gerhard  
TITLE OF INVENTION: CATHEPSIN D IS AN AMYLOIDGENIC  
TITLE OF INVENTION: PROTEASE IN ALZHEIMER S DISEASE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Miles Inc.  
STREET: 400 Morgan Lane  
CITY: West Haven  
STATE: Connecticut  
COUNTRY: USA  
ZIP: 06516  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
COMPUTER: Sharp PC 4600  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/07043A  
FILING DATE: June 21, 1994



CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10889  
FILING DATE: November 12, 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/995,660  
FILING DATE: December 16, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/880,914  
FILING DATE: May 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pamela A. Simonton  
REGISTRATION NUMBER: 31,060  
REFERENCE/DOCKET NUMBER: MTI 224.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (203) 937-2340  
TELEFAX: (203) 937-2795  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US94-07043A-3

Query Match 84.2% Score 32; DB 5; Length 11;  
Best Local Similarity 87.5% Pred. No. 0.14;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAEF 8  
1111111  
Db 3 EVNLDAEF 10

RESULT 5  
US-08-659-984A-18  
Sequence 18, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-0028100S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-18

Query Match 84.2% Score 32; DB 2; Length 21;  
Best Local Similarity 87.5% Pred. No. 0.29;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAEF 8  
1111111  
Db 2 EVNLDAEF 9

RESULT 6  
US-08-802-981-112  
Sequence 112, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:  
APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly S.  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-000300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 112:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product="A1b"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /product="Acp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 16  
OTHER INFORMATION: /product="Acp"  
US-08-802-981-112  
Query Match 84.2% Score 32; DB 3; Length 21;

Best Local Similarity 87.5%; Pred. No. 0.29;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVNLAEF 8  
|||||  
Db 7 EVNLDAEF 14

RESULT 7  
US-08-660-531-18

Sequence 18, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-18

Query Match 84.2%; Score 32; DB 4; Length 21;  
Best Local Similarity 87.5%; Pred. No. 0.29;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVNLAEF 8  
|||||  
Db 2 EVNLDAEF 9

RESULT 8  
US-08-659-984A-17

Sequence 17, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Crook, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase

TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422

INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-659-984A-17

Query Match 84.2%; Score 32; DB 2; Length 30;  
Best Local Similarity 87.5%; Pred. No. 0.45;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVNLAEF 8  
|||||  
Db 23 EVNLDAEF 30

RESULT 9

US-08-660-531-17

Sequence 17, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-17

Query Match 84.2%; Score 32; DB 4; Length 30;  
Best Local Similarity 87.5%; Pred. No. 0.45;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAEP 8  
DB 23 EVNLAEP 30

RESULT 10  
US-08-659-984A-16  
Sequence 16, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-08-659-984A-16

Query Match 84.2%; Score 32; DB 2; Length 33;  
Best Local Similarity 87.5%; Pred. No. 0.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAEP 8  
DB 14 EVNLAEP 21

RESULT 11  
US-08-660-531-16  
Sequence 16, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-16

Query Match 84.2%; Score 32; DB 4; Length 33;  
Best Local Similarity 87.5%; Pred. No. 0.5;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAEP 8  
DB 14 EVNLAEP 21

RESULT 12  
US-08-659-984A-15  
Sequence 15, Application US/08659984A  
Patent No. 5942400

GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-15

Query Match 84.2%; Score 32; DB 2; Length 42;  
Best Local Similarity 87.5%; Pred. No. 0.66;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVNLAEF 8  
1111111  
DB 23 EVNLDAEF 30

RESULT 13  
US-08-660-531-15  
Sequence 15, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-15

Query Match 84.2%; Score 32; DB 4; Length 42;  
Best Local Similarity 87.5%; Pred. No. 0.66;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVNLAEF 8  
1111111  
DB 23 EVNLDAEF 30

RESULT 14  
US-08-339-708A-12  
Sequence 12, Application US/08339708A  
Patent No. 6037521  
GENERAL INFORMATION:  
APPLICANT: Sato, Masahiro  
APPLICANT: Takashi, Kobayashi  
APPLICANT: Shoji, Mikio  
APPLICANT: Kawarabayashi, Takeshi  
TITLE OF INVENTION: TRANSGENIC ANIMAL MODEL FOR ALZHEIMER'S  
TITLE OF INVENTION: DISEASE  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/339,708A  
FILING DATE: 14-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 306026/93  
FILING DATE: 12-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SANDERCOCK, COLIN G.  
REGISTRATION NUMBER: 31,298  
REFERENCE/DOCKET NUMBER: 026083/0159  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399

TELEX: 904136  
INFORMATION FOR SEQ ID NO: 12  
SEQUENCE CHARACTERISTICS:  
LENGTH: 103 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-339-708A-12

Search completed: September 6, 2001, 16:39:39  
Job time: 128 sec

Query Match 84.2%; Score 32; DB 3; Length 103;  
Best Local Similarity 87.5%; Pred. No. 1.9;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
1111111  
DB 1 EVNLDAEF 8

RESULT 15  
US-08-659-984A-21  
Sequence 21, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 506 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-659-984A-21

Query Match 84.2%; Score 32; DB 2; Length 506;  
Best Local Similarity 87.5%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
1111111  
DB 404 EVNLDAEF 411

Fri Sep 7 10:58:11 2001

us-09-603-713-28.ra1

Page 8

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:32 ; Search time 225.25 Seconds  
(without alignments)  
2.153 Million cell updates/sec

Title: US-09-603-713-28  
Perfect score: 38  
Sequence: 1 EVNLAAPF 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues  
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0601.\*  
1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	8	22	AAB66586 Human memapsin 2 1
2	38	100.0	8	22	AAB61348 Peptide OM99-2. U
3	38	100.0	9	21	AAB07879 A peptide fragment
4	34	89.5	9	21	AAB07885 A peptide fragment
5	34	89.5	14	21	AAB07888 A peptide fragment
6	33	86.8	7	22	AAB65585 Human memapsin 2 1
7	33	86.8	7	22	AAB61347 Peptide OM99-1. U
8	33	86.8	9	21	AAB07882 A peptide fragment
9	32	84.2	8	21	AAV94771 Beta-secretase sub
10	32	84.2	9	19	AAW82081 Fluorogenic protea
11	32	84.2	9	21	AAB07874 A peptide fragment

12	32	84.2	9	21	AAB07880 A peptide fragment
13	32	84.2	9	21	AAB07894 Substrate for beta
14	32	84.2	10	18	AAM08362 Beta-secretase sub
15	32	84.2	10	20	AAV33756 Synthetic oligopep
16	32	84.2	10	21	AAV69707 Beta-APP alpha-sec
17	32	84.2	10	22	SYB66575 Synthetic peptide
18	32	84.2	10	22	AAB61337 Swedish mutation p
19	32	84.2	16	21	AAB06316 Human beta-amyloid
20	32	84.2	20	19	AAW82211 Fluorogenic protea
21	32	84.2	20	21	AAV69714 Beta-APP alpha-sec
22	32	84.2	21	18	AAM08361 Beta-secretase sub
23	32	84.2	21	19	AAW82186 Fluorogenic protea
24	32	84.2	21	20	AAV33755 Synthetic oligopep
25	32	84.2	30	18	AAW08360 Beta-secretase sub
26	32	84.2	30	20	AAV33754 Synthetic oligopep
27	32	84.2	30	21	AAB07895 Substrate for beta
28	32	84.2	32	17	AAW04402 Mouse amyloid prec
29	32	84.2	32	17	AAW04403 Mouse amyloid prec
30	32	84.2	32	17	AAW04401 Mouse amyloid prec
31	32	84.2	33	18	AAW08359 Beta-secretase sub
32	32	84.2	33	20	AAV33753 Synthetic oligopep
33	32	84.2	33	21	AAB07892 Substrate for beta
34	32	84.2	39	21	AAV69718 Beta-APP alpha-sec
35	32	84.2	42	18	AAW08350 Wild-type APP beta
36	32	84.2	42	20	AAV33752 Synthetic oligopep
37	32	84.2	58	20	AAW98001 Swedish-FAD APP714
38	32	84.2	103	16	AAW74698 Beta-amyloid precu
39	32	84.2	115	20	AAW97997 Swedish-FAD APP po
40	32	84.2	115	20	AAW98000 Swedishshlonon-FAD
41	32	84.2	506	19	AAW61152 Maltose binding pr
42	32	84.2	506	20	AAV33742 MBP-APP (SW192) fu
43	32	84.2	595	21	AAV88435 Human APP695-sw va
44	32	84.2	697	21	AAV88429 Human APPSW-KK aml
45	31	81.6	106	21	AAV75143 Nisseria meningit

## ALIGNMENTS

RESULT 1		AAB66586 standard: Peptide; 8 AA.	
ID	AAB66586		
XX	AAAB66586;		
AC	12-APR-2001 (first entry)		
XX			
DT	Human memapsin 2 inhibitor OM99-2.		
XX			
DE			
XX			
KW	Human memapsin 2; neurotrophic; neuroprotective; amyloid precursor protein;		
KW	APP; memapsin 2 inhibitor; Alzheimer's disease.		
XX			
OS	Synthetic.		
XX			
XX	Key	Location/Qualifiers	
FT	Modified-site	4.5	
FT	/note="residues 4 and 5 form a Leu-Ala dipeptide		
FT	isotere"		
XX			
PN	W0200100665-A2.		
XX			
PD	04-JAN-2001.		
XX			
XX	27-JUN-2000; 2000WO-US17742.		
XX			
PR	28-JUN-1999; 99US-0141363.		
PR	30-NOV-1999; 99US-0168060.		
PR	25-JAN-2000; 2000US-0177836.		
PR	27-JAN-2000; 2000US-0178368.		
PR	08-JUN-2000; 2000US-0210292.		
XX			
PA	(OKLA-) OKLAHOMA MEDICAL RES FOUND.		
PA	(UNIT ) UNIV ILLINOIS FOUND.		

XX Tang JUN, Hong L, Ghosh AK;  
 PI WPI: 2001-137933/14.  
 DR  
 XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage -  
 XX  
 XX Example 7; Page 36; 86pp; English.  
 PS  
 CC The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 CC  
 XX Sequence 8 AA;  
 SQ

Query Match 100.0%; Score 38; DB 22; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 |||||  
 Db 1 evnlaaef 8

## RESULT 2

AAB61348  
 ID AAB61348 standard; peptide; 8 AA.  
 XX

AC AAB61348;  
 XX

DT 02-APR-2001 (first entry)  
 XX

DE Peptide OM99-2.  
 XX

KM Memapsin 2; catalyst; Alzheimer's.  
 XX

OS Unidentified.  
 XX

PN WO200100663-A2.  
 XX

PD 04-JAN-2001.  
 XX

PF 27-JUN-2000; 2000WO-US17661.  
 XX

PR 28-JUN-1999; 99US-0141363.  
 XX

PR 30-NOV-1999; 99US-0168060.  
 XX

PR 25-JAN-2000; 2000US-0177836.  
 XX

PR 27-JAN-2000; 2000US-0178368.  
 XX

PR 08-JUN-2000; 2000US-0210292.  
 XX

PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 XX

PI Tang JUN, Lin X, Koelsch G;  
 XX

DR WPI: 2001-102885/11.  
 XX

XX Purified recombinant catalytically active memapsin 2, used to screen  
 PT inhibitors of it, which are used to treat and prevent Alzheimer's  
 PT disease -  
 XX

PS Example 7; Page 36; 86pp; English.  
 XX The present invention relates to a purified recombinant  
 CC

CC catalytically active memapsin 2. The invention may be used for  
 CC isolating inhibitors which are used to treat or prevent  
 CC Alzheimer's disease. The invention may also be used to screen  
 CC for individuals more genetically prone to develop Alzheimer's  
 CC disease.  
 CC  
 XX Sequence 8 AA;  
 SQ

Query Match 100.0%; Score 38; DB 22; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 |||||  
 Db 1 evnlaaef 8

## RESULT 3

AAB07879  
 ID AAB07879 standard; peptide; 9 AA.  
 XX

AC AAB07879;  
 XX

DT 14-NOV-2000 (first entry)  
 XX

DE A peptide fragment derived from beta-amyloid precursor protein.  
 XX

KM Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KM amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 KM inhibitor.  
 XX

KW Homo sapiens.  
 XX

OS WO200047618-A2.  
 XX

PN 17-AUG-2000.  
 XX

PD 10-FEB-2000; 2000WO-US03819.  
 XX

PF 10-FEB-1999; 99US-0119571.  
 XX

PR 15-JUN-1999; 99US-0139172.  
 XX

XX (ELAN-) ELAN PHARM INC.  
 PA

XX Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
 XX

DR WPI: 2000-533011/48.  
 XX

XX Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease -  
 XX

PS Disclosure; Page 12; 121pp; English.  
 XX

XX The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents a peptide derived from beta-amyloid  
 CC precursor protein  
 XX

SQ Sequence 9 AA;  
 XX

Query Match 100.0%; Score 38; DB 21; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;



Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 |||||  
 Db 2 evnlaaef 9

## RESULT 4

AAB07885  
 ID AAB07885 standard; peptide; 9 AA.

AC AAB07885;  
 XX

DT 14-NOV-2000 (first entry)  
 XX

DE A peptide fragment derived from beta-amyloid precursor protein.  
 XX

KM Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 XX inhibitor.  
 XX

OS Homo sapiens.  
 XX

PN WO20047618-A2.  
 XX

PD 17-AUG-2000.  
 XX

PF 10-FEB-2000; 2000WO-US03819.  
 XX

PR 10-FEB-1999; 99US-0119571.  
 PR 15-JUN-1999; 99US-0139172.  
 XX

PA (ELAN-) ELAN PHARM INC.  
 XX

PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
 XX

DR WPI; 2000-533011/48.  
 XX

PT Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 XX Alzheimer's disease -  
 XX

PS Disclosure: Page 12; 121pp; English.  
 XX

CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents a peptide derived from beta-amyloid  
 CC precursor protein  
 CC  
 XX

SO Sequence 9 AA;

Query Match 89.5%; Score 34; DB 21; Length 9;  
 Best Local Similarity 87.5%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 |||||  
 Db 2 evnlaaef 9

## RESULT 5

AAB07888  
 ID AAB07888 standard; peptide; 14 AA.  
 XX

AC AAB07888;  
 XX

DT 14-NOV-2000 (first entry)  
 XX

DE A peptide fragment derived from beta-amyloid precursor protein.  
 XX  
 KM Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 XX inhibitor.  
 XX

OS Homo sapiens.  
 XX

PN WO20047618-A2.  
 XX

PD 17-AUG-2000.  
 XX

PF 10-FEB-2000; 2000WO-US03819.  
 XX

PR 10-FEB-1999; 99US-0119571.  
 PR 15-JUN-1999; 99US-0139172.  
 XX

PA (ELAN-) ELAN PHARM INC.  
 XX

PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
 XX

DR WPI; 2000-533011/48.  
 XX

PT Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 XX Alzheimer's disease -  
 XX

PS Disclosure: Page 12; 121pp; English.  
 XX

CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents a peptide derived from beta-amyloid  
 CC precursor protein  
 CC  
 XX

SO Sequence 14 AA;

Query Match 89.5%; Score 34; DB 21; Length 14;  
 Best Local Similarity 87.5%; Pred. No. 0.19;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNLAAEF 8  
 |||||  
 Db 7 evnlvaef 14

## RESULT 6

AAB66585  
 ID AAB66585 standard; Peptide; 7 AA.

AC AAB66585;  
 XX

DT 12-APR-2001 (first entry)  
 XX

DE Human memapsin 2 inhibitor OM99-1.  
 XX

KM Human memapsin 2; nootropic; neuroprotective; amyloid precursor protein;  
 KW APP; memapsin 2 inhibitor; Alzheimer's disease.  
 XX  
 XX Synthetic.  
 OS

FH Key Location/Qualifiers  
 FT Modified-site 3..4  
 FT /note= "residues 3 and 4 form a Leu-Ala dipeptide  
 FT Isotere"  
 XX  
 XX  
 PN WO200100665-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 27-JUN-2000; 2000WO-US17742.  
 XX  
 PR 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX  
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA (UNIT ) UNIV ILLINOIS FOUND.  
 XX  
 PI Tang JUN, Hong L, Ghosh AK;  
 PI  
 PI WPI: 2001-137933/14.  
 DR  
 XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage -  
 PT  
 XX Example 7; Page 36; 86pp; English.  
 PS  
 XX The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 CC  
 CC  
 CC Sequence 7 AA:  
 XX  
 SQ  
 Query Match 86.8%; Score 33; DB 22; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 VNLAEEF 8  
 DB 1 vnlaaef 7  
 DE  
 RESULT 7  
 AAB61347  
 ID AAB61347 standard; peptide; 7 AA.  
 XX  
 AC AAB61347;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Peptide OM99-1.  
 XX  
 KW Memapsin 2; catalyst; Alzheimer's.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200100663-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 27-JUN-2000; 2000WO-US17661.  
 XX

PR 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX  
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA  
 PI Tang JUN, Lin X, Koelsch G;  
 PI  
 PI WPI: 2001-102885/11.  
 DR  
 XX Purified recombinant catalytically active memapsin 2, used to screen  
 PT inhibitors of it, which are used to treat and prevent Alzheimer's  
 PT disease -  
 PT  
 XX Example 7; Page 36; 86pp; English.  
 PS  
 XX The present invention relates to a purified recombinant  
 CC catalytically active memapsin 2. The invention may be used for  
 CC isolating inhibitors which are used to treat or prevent  
 CC Alzheimer's disease. The invention may also be used to screen  
 CC for individuals more genetically prone to develop Alzheimer's  
 CC disease.  
 CC  
 CC  
 CC Sequence 7 AA:  
 XX  
 SQ  
 Query Match 86.8%; Score 33; DB 22; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 VNLAEEF 8  
 DB 1 vnlaaef 7  
 DE  
 RESULT 8  
 AAB07882  
 ID AAB07882 standard; peptide; 9 AA.  
 XX  
 AC AAB07882;  
 XX  
 DT 14-NOV-2000 (first entry)  
 XX  
 DE A peptide fragment derived from beta-amyloid precursor protein.  
 XX  
 KW beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 KW inhibitor.  
 KW  
 KW Homo sapiens.  
 OS  
 PN WO200047618-A2.  
 PN  
 PD 17-AUG-2000.  
 XX  
 PF 10-FEB-2000; 2000WO-US03819.  
 XX  
 PR 10-FEB-1999; 99US-0119571.  
 PR 15-JUN-1999; 99US-0139172.  
 XX  
 PA (ELAN-) ELAN PHARM INC.  
 PA  
 PI Anderson JP, Basi G, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
 PI  
 PI WPI: 2000-533011/48.  
 DR  
 XX Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease -  
 PT  
 XX

PS Disclosure: Page 12; 121pp; English.

CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents a peptide derived from beta-amyloid  
 CC precursor protein

CC Sequence 9 AA:

XX

QY 1 EVNLAEF 8  
 |||||  
 Db 2 evnlaeef 9

RESULT 9

AA94771  
 ID AAY94771 standard; Protein; 8 AA.

AC AAY94771;  
 XX

DT 12-FEB-2001 (first entry)  
 XX

DE Beta-secretase substrate peptide SEQ ID 17.  
 XX

KM Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease;  
 XX

KW Down's syndrome; amyloid angiodopathy; gene therapy; neuroprotective.  
 XX

OS Synthetic.  
 XX

PN WO200058479-A1.  
 XX

PD 05-OCT-2000.  
 XX

PF 23-MAR-2000: 2000MO-US07755.  
 XX

PR 26-MAR-1999: 99US-0277229.  
 XX

PA (AMGE-) AMGEN INC.  
 XX

PI Citron M, Vassar RJ, Bennett BD;  
 XX

DR WPI: 2000-594643/56.  
 XX

PT Isolated beta-secretase nucleic acids and encoded polypeptides, useful  
 XX

PT for diagnosis and gene therapy of Alzheimer's disease -  
 XX

PS Example 10; Page 117; 145pp; English.

XX

CC This invention relates to 3 nucleotide sequences encoding beta-secretase  
 CC proteins. Beta-secretase is an enzyme involved in the production of one  
 CC of the components of amyloid plaques involved in Alzheimer's disease. The  
 CC invention includes an expression vector comprising the nucleotide  
 CC sequence, a host cell comprising the expression vector, and a process for  
 CC producing the protein through culturing the transformed cells. Also  
 CC included in the invention are a polypeptide derivative of the  
 CC beta-secretase protein, a fusion protein comprising beta-secretase fused  
 CC to a heterologous amino acid sequence, and a method for modulating the  
 CC levels of beta-secretase polypeptide in a mammal comprising administering  
 CC the polynucleotide sequence. Beta-secretase exhibits neuroprotective and  
 CC neurotrophic activity. The beta-secretase nucleotide sequence may be used to  
 CC map locations of the beta-secretase gene and related genes on chromosomes

CC and as hybridization probes in diagnostic assays to test for the presence  
 CC of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's  
 CC syndrome, and amyloid angiodopathy. The nucleotide sequence may also be  
 CC used as anti-sense inhibitors of beta-secretase expression. In gene  
 CC therapy of Alzheimer's disease, and for the identification of compounds  
 CC that modulate beta-secretase activity. Antibodies to the beta-secretase  
 CC protein may be used for in vitro and in vivo diagnostic purposes to  
 CC detect the presence of beta-secretase polypeptide in a body fluid or cell  
 CC sample. The present sequence represents a beta-secretase substrate  
 CC peptide.

CC Sequence 8 AA:

XX

QY 1 EVNLAEF 8  
 |||||  
 Db 1 evnlaeef 8

RESULT 10

AAW82081  
 ID AAW82081 standard; peptide; 9 AA.

AC AAW82081;  
 XX

DT 18-FEB-1999 (first entry)  
 XX

DE Fluorogenic protease indicator protease binding peptide #59.  
 XX

KM Protease activity; fluorophore; detection; fluorogenic; cellular uptake;  
 XX

KW conformation change.  
 XX

OS Synthetic.  
 XX

PN WO9837226-A1.  
 XX

PD 27-AUG-1998.  
 XX

PF 20-FEB-1998: 98MO-US03000.  
 XX

PR 20-FEB-1997: 97US-0802981.  
 XX

PA (ONCO-) ONCOIMMUNIN INC.  
 XX

PI Komoriya A, Packard BS;  
 XX

DR WPI: 1998-46579/40.  
 XX

PT New fluorogenic compositions - containing 2 fluorophores separated  
 XX

PT by a peptide comprising a protease binding site, used for detecting  
 XX

PT protease activity in samples.  
 XX

PS Claim 4; Page 77; 90pp; English.

XX

CC AAW82023-W82240 are peptides used in the construction of a fluorogenic  
 CC composition which is used for the detection of protease activity in  
 CC biological samples. The products can be used for the detection of  
 CC conformational changes in nucleic acids, oligosaccharides,  
 CC polysaccharides, proteins, peptides, lipids, phospholipids, glycolipids,  
 CC glycoproteins, steroids or polymers. In addition, attachment of a  
 CC hydrophobic group to a molecule can be used to enhance uptake by cells.  
 CC The composition is composed of P = peptide comprising a protease binding  
 CC site for the protease, F1, F2 peptides = fluorophores where F1 is  
 CC attached to the amino terminal amino acid and F2 is attached to the  
 CC carboxyl terminal amino acid and S1, S2 peptides = when present, are  
 CC peptide spacers where S1, when present, is attached to the amino terminal  
 CC acid, and S2, when present, is attached to the carboxyl terminal amino  
 CC acid.

SQ Sequence 9 AA;

Query Match 84.2%; Score 32; DB 19; Length 9;  
 Best Local Similarity 87.5%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNIAAEF 8  
 |||||  
 DB 2 evnladeef 9

RESULT 11  
 AAB07874  
 ID AAB07874 standard; peptide: 9 AA.

AC AAB07874;

DT 14-NOV-2000 (first entry)

DE A peptide fragment derived from beta-amyloid precursor protein.

XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 KW inhibitor.

OS Homo sapiens.

XX MO200047618-A2.

XX 17-AUG-2000.

PP 10-FEB-2000; 2000MO-US03819.

PR 10-FEB-1999; 9905-0119571.

PR 15-JUN-1999; 9905-0139172.

XX (ELAN-) ELAN PHARM INC.

XX Anderson JP, Basl G, Doane MT, Frigon N, John V, Power M;

PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

DR WPI: 2000-533011/48.

PP Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease.

PS Disclosure; Page 12; 121pp; English.

XX The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents a peptide derived from beta-amyloid  
 CC precursor protein  
 XX  
 SQ Sequence 9 AA;

Query Match 84.2%; Score 32; DB 21; Length 9;  
 Best Local Similarity 87.5%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNIAAEF 8  
 |||||  
 DB 2 evnladeef 9

RESULT 12

ID AAB07880 standard; peptide: 9 AA.

AC AAB07880;

DT 14-NOV-2000 (first entry)

DE A peptide fragment derived from beta-amyloid precursor protein.

XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 KW inhibitor.

OS Homo sapiens.

XX MO200047618-A2.

XX 17-AUG-2000.

PP 10-FEB-2000; 2000MO-US03819.

PR 10-FEB-1999; 9905-0119571.

PR 15-JUN-1999; 9905-0139172.

XX (ELAN-) ELAN PHARM INC.

XX Anderson JP, Basl G, Doane MT, Frigon N, John V, Power M;

PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

DR WPI: 2000-533011/48.

PP Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease.

PS Disclosure; Page 12; 121pp; English.

XX The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents a peptide derived from beta-amyloid  
 CC precursor protein  
 XX  
 SQ Sequence 9 AA;

Query Match 84.2%; Score 32; DB 21; Length 9;  
 Best Local Similarity 87.5%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EVNIAAEF 8  
 |||||  
 DB 2 evnladeef 9

RESULT 13

ID AAB07894 standard; Peptide: 9 AA.

AC AAB07894;

DT 14-NOV-2000 (first entry)

DE Substrate for beta-secretase enzyme.

OS	Synthetic.		
XX			
FH	Key	Location/Qualifiers	
FT	Cleavage-site	5..6	
XX			
PN	MO200047618-A2.		
PD			
XX	17-AUG-2000.		
XX			
PF	10-FEB-2000; 2000MO-USO3819.		
XX			
PR	10-FEB-1999; 99US-0119571.		
PR	15-JUN-1999; 99US-0139172.		
XX			
PA	(ELAN-) ELAN PHARM INC.		
PI	Anderson JP, Basil G, Doane MT, Frigon N, John V, Power M;		
PI	Sluha S, Tatsuno G, Tung J, Wang S, McConlogue L;		
XX	WPI: 2000-533011/48.		
DR			
XX			
PT	Purified beta-secretase protein used in assays to discover inhibitors		
PT	which can be used for the treatment of amyloidogenic diseases e.g.		
PT	Alzheimer's disease -		
XX			
PS	Example 4; Page 71; 121pp; English.		
XX			
CC	The specification describes a beta-secretase enzyme. The enzyme cleaves		
CC	beta-amyloid precursor protein to produce beta-amyloid peptide. This		
CC	enzyme is therefore implicated in the production of amyloid plaque		
CC	components which accumulate in the brains of individuals afflicted with		
CC	Alzheimer's disease. Inhibitors of beta-secretase are administered to		
CC	a mammalian subject e.g. with Alzheimer's disease or Alzheimer's		
CC	disease-like pathology to test if they maintain or improve cognitive		
CC	ability or reduce the plaque burden. The compounds are used for the		
CC	treatment of amyloidogenic diseases e.g. Alzheimer's disease. The		
CC	present sequence represents a peptide substrate used to test the		
CC	activity of beta-secretase enzyme.		
XX			
CC			
SO	Sequence	9 AA:	
	Query Match	84.2%; Score 32; DB 21; Length 9;	
	Best Local Similarity	87.5%; Pred. NO. 3.4e+05;	
	Matches	7; Conservative	0; Mismatches
		1; Indels	0; Gaps
OY			
	1 EVN1AEP	8	
DB	2 evn1daef	9	
	RESULT	14	
	AAW08362		
ID	AAW08362 standard; peptide; 10 AA.		
AC	AAW08362;		
XX			
XX	05-SEP-1997 (first entry)		
XX			
DE	Beta-secretase substrate #3.		
XX			
KM	Beta-cleavage site; beta amyloid precursor protein; APP; beta-secretase;		
KM	alpha-secretase; proteolytic cleavage; inhibitor; Alzheimer's disease.		
XX			
OS	Synthetic.		
XX			
FH	Key	Location/Qualifiers	
FT	Misc-difference	1	
FT		/note="acetylated"	

XX	PN	MO9640885-A2.	
XX	PD	19-DEC-1996.	
XX	PE	07-JUN-1996;	96WO-US09985.
XX	PR	07-JUN-1995;	95US-0485152.
XX	PP	07-JUN-1995;	95US-0480498.
XX	PA	(ATHE-) ATHENA NEUROSCIENCES INC.	
XX	PI	Anderson JP, Chrysler SMS, Jacobson-croak KL, Kelin PS;	
XX	PI	Mcconlogue LC, Sinha S, Tan H;	
XX	DR	WPI: 1997-052304/05.	
XX	PT	Beta-secretase which specifically cleaves beta-amyloid precursor	
XX	PT	protein - useful to screen for inhibitors useful in treatment of	
XX	PT	Alzheimer's disease	
XX	PS	Disclosure: Page 45; 92pp; English.	
XX	XX	AAW08359-W08362, represent substrates for the enzyme of the invention.	
XX	CC	The enzyme of the invention is beta-secretase, and specifically cleaves	
XX	CC	beta-amyloid precursor protein (beta-APP). Normal processing of beta-APP	
XX	CC	is thought to occur via cleavage between residues 16 and 17 of the	
XX	CC	beta-amyloid peptide region by an alpha-secretase. Pathogenic processing	
XX	CC	is thought to occur by beta-secretase cleavage of beta-APP.	
XX	CC	Beta-secretase activity can be detected and measured using a method of	
XX	CC	the invention, which detects at least one of the beta-secretase cleavage	
XX	CC	products formed on cleavage. The method can be used to determine whether	
XX	CC	a test substance inhibits proteolytic cleavage, by beta-secretase, of	
XX	CC	beta-APP. Compounds effective to at least partially inhibit	
XX	CC	beta-secretase activity can be used to inhibit cleavage of beta-APP in	
XX	CC	cells or mammalian hosts. Isolation and purification of beta-secretase	
XX	CC	will permit chemical modelling of a critical event in the pathology of	
XX	CC	Alzheimer's disease.	
XX	SQ	Sequence 10 AA:	
XX	QY	Query Match 84.2%; Score 32; DB 18; Length 10;	
XX	QY	Best Local Similarity 87.5%; Pred. No. 0.37;	
XX	QY	Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
XX	QY	1 EYNLAAPF 8	
XX	QY		
XX	QY	2 EVNIDAEF 9	
XX	DE	RESULT 15	
XX	ID	AAV33756	
XX	AC	AAV33756 standard; Protein; 10 AA.	
XX	AC	AAV33756;	
XX	DT	09-NOV-1999 (first entry)	
XX	DE	Synthetic oligopeptide 5-5'SW.	
XX	KM	Beta-secretase; beta-amyloid protein precursor; APP; Down's syndrome;	
XX	KW	Alzheimer's disease; measure activity; cleavage site.	
XX	OS	Synthetic.	
XX	OS	Key	Location/Qualifiers
XX	FT	Modified-site 1	
XX	FT	/note- "N-terminal Ser is acetylated"	
XX	PN	US5942400-A.	
XX	PD	24-AUG-1999.	

```

XX      07-JUN-1996;      96US-0659984.
PF      07-JUN-1996;      96US-0659984.
XX      07-JUN-1995;      95US-0480498.
PR      07-JUN-1995;      95US-0480498.
FR      07-JUN-1995;      95US-0485152.
XX      (ELAN-) ELAN PHARM INC.
PA
XX      Anderson JP, Jacobson-Croak KL, Sinha S;
PI      WPI: 1999-517417/43.
DR
XX      A method for detecting human beta-secretase cleavage of polypeptides
PT      useful for identifying beta-secretase inhibitors
XX
XX      Examples; Column 30; 43pp; English.
RS
XX      Sequences AAY3752-Y3756 are synthetic oligopeptides used for measuring
CC      the activity of beta-secretase (AAY3741). Beta-secretase is capable of
CC      cleaving beta-amyloid protein precursor (APP) (AAY3742). These
CC      synthetic peptides contain the cleavage site of APP. Beta-secretase and
CC      APP are used in a method for detecting human beta-secretase cleavage of
CC      polypeptides and for identifying beta-secretase inhibitors. Inhibition
CC      of beta-secretase activity would be useful for chemical modeling of a
CC      critical event in the pathology of Alzheimer's disease. Inhibitors of
CC      beta-secretase would be useful for the prevention and treatment of
CC      Alzheimer's disease and Down's Syndrome.
XX
SQ      Sequence      10 AA;

```

```

Query Match      84.2%; Score 32; DB 20; Length 10;
Best Local Similarity 87.5%; Pred. No. 0.37;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 EVNLAEEF 8
      ||| |||
Db      2 evndaef 9

```

Search completed: September 6, 2001, 16:43:32  
Job time: 361 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:55 ; Search time 231.42 Seconds  
(without alignments)  
4.002 Million cell updates/sec

Title: US-09-603-713-27

Perfect score: 33

Sequence: 1 VNLAEF 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_16:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mmc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_protent:\*

12: sp\_unclassified:\*

13: sp\_vertebrate:\*

14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	87.9	328	2 066974	066974 aquifex aeo
2	29	87.9	657	3 090003	090003 orpiniomycs
3	29	87.9	797	5 096V13	096V13 hydra magni
4	29	87.9	2270	14 09JFN3	09JFN3 lupala para
5	28	84.8	245	10 09LDV6	09LDV6 oryza sativ
6	28	84.8	280	1 09HNT6	09HNT6 halobacteri
7	28	84.8	315	2 P94419	P94419 bacillus su
8	28	84.8	629	3 09P4E1	09P4E1 cunningham
9	28	84.8	888	3 09HEP0	09HEP0 candida alb
10	28	84.8	906	5 017281	017281 botryllus s
11	28	84.8	1283	5 09W215	09W215 drosophila
12	28	84.8	1332	5 09YEC4	09YEC4 drosophila
13	27	81.8	123	10 09FKH5	09FKH5 arabidopsis
14	27	81.8	223	2 085605	085605 prevotella
15	27	81.8	251	5 P90781	P90781 caenorhabd1
16	27	81.8	282	2 030485	030485 streptomyce
17	27	81.8	291	2 09XE10	09XE10 enterobacte
18	27	81.8	291	2 09RIR4	09RIR4 serratia fo
19	27	81.8	291	2 09R1Q9	09R1Q9 serratia fo

20	27	81.8	291	2 09R1Q7	09R1Q7 serratia fo
21	27	81.8	291	2 09R444	09R444 serratia fo
22	27	81.8	295	2 046990	046990 enterobacte
23	27	81.8	317	2 09KK63	09KK63 streptomyce
24	27	81.8	335	3 09P6L3	09P6L3 schizosacch
25	27	81.8	370	4 09MMO2	09MMO2 homo saplen
26	27	81.8	385	2 046696	046696 lactococcus
27	27	81.8	388	2 046697	046697 lactococcus
28	27	81.8	388	2 054685	054685 lactococcus
29	27	81.8	401	2 048702	048702 lactococcus
30	27	81.8	401	2 054672	054672 lactococcus
31	27	81.8	493	4 015323	015323 homo saplen
32	27	81.8	500	3 094330	094330 schizosacch
33	27	81.8	538	10 09FZH1	09FZH1 arabidopsis
34	27	81.8	606	2 059675	059675 pseudomonas
35	27	81.8	643	2 092663	092663 vibrio para
36	27	81.8	790	1 027578	027578 methanobact
37	26	78.8	108	5 09GWD1	09GWD1 leishmania
38	26	78.8	131	14 091111	091111 human immun
39	26	78.8	225	14 084614	084614 parametium
40	26	78.8	241	3 012580	012580 chaetomium
41	26	78.8	282	5 018230	018230 caenorhabd1
42	26	78.8	293	4 060812	060812 homo saplen
43	26	78.8	295	10 09SHF7	09SHF7 arabidopsis
44	26	78.8	300	1 028459	028459 archaeoglob
45	26	78.8	306	6 077768	077768 oryctolagus

## ALIGNMENTS

RESULT 1

ID 066974 PRELIMINARY: PRT: 328 AA.

AC 066974:

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE HYPOTHETICAL 36.0 KDA PROTEIN.

GN AO.778.

GN Aquifex aeolicus.

OC Bacteria; Aquificales; Aquificaceae; Aquifex.

OX NCBI\_TaxID=63363;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VF5.

RX MEDLINE=98196666; PubMed=9537320;

RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,

RA Graham D.E., Overbeek R., Sneed M.A., Keller M., AuJay M., Huber R.,

RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;

RT "The complete genome of the hyperthermophilic bacterium Aquifex

RT aeolicus.";

RT Nature 392:353-358(1998).

DR EMBL: AE000707; AAC06942.1; -.

DR InterPro: IPR000873; -.

DR InterPro: IPR002882; -.

DR Pfam: PF01933; UPF0052; 1.

DR PROSITE: PS00455; AMP\_BINDING; UNKNOWN\_1.

KW Hypothetical protein.

SO SEQUENCE 328 AA; 35957 MW; 30BB214E9950F6B CRC64;

Query Match 87.9%; Score 29; DB 2; Length 328;

Best local Similarity 85.7%; Pred. No. 57;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VNLAEF 7

DB 133 VNLAEF 139

RESULT 2

Q90003

ID 09UN03 PRELIMINARY: PRT: 657 AA.  
 AC 09UN03:  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE BETA-GLUCOSIDASE (EC 3.2.1.21).  
 GN BGL1.  
 OS Orpinomyces sp. PC-2.  
 OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;  
 OC Neocallimastixaceae; Orpinomyces.  
 OX NCBI\_TaxID=50059;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PC-2;  
 RA "Kliment E.A., Chen H., Li X., Felix C.R., Ljungdahl L.G.;  
 "Cloning and sequencing of a beta-glucosidase from the anaerobic  
 fungus Orpinomyces sp. strain PC-2";  
 RT Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF016864; AAD45834.1; -.  
 DR HSSP: P26205; ICBG.  
 DR InterPro: IPR001360; -.  
 DR InterPro: IPR002667; -.  
 DR Pfam: PF00232; Glyco\_hydro\_1; 3.  
 DR PRINTS: PR00131; GLHYDRLASE1.  
 DR Prodom: PD004109; -; 1.  
 DR PROSITE: PS00653; GLYCOSYL\_HYDROL\_FL\_2; 1.  
 KW Hydrolase; Glycosidase.  
 SQ SEQUENCE 657 AA; 75227 MW; C228E706C037478B CRC64;

Query Match 87.9%; Score 29; DB 3; Length 657;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 NLAEEF 7  
 |||||  
 DB 377 NLAEEF 382  
 RESULT 3  
 O9GV13 PRELIMINARY: PRT: 797 AA.  
 AC 09GV13:  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE VASA-RELATED PROTEIN CNVAS1.  
 GN CNVAS1.  
 OS Hydra magnipapillata (Hydra).  
 OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Anthomedusae;  
 OC Hydridae; Hydra.  
 OX NCBI\_TaxID=6085;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mochizuki K., Fujisawa T.;  
 "Vasa-related gene in hydra";  
 RT Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AB047382; BAB13307.1; -.  
 SQ SEQUENCE 797 AA; 83935 MW; 847D939DA1AF622C CRC64;

Query Match 87.9%; Score 29; DB 5; Length 797;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 NLAEEF 7  
 |||||  
 DB 556 NLAEEF 561  
 RESULT 4  
 O9JEN3 PRELIMINARY: PRT: 2270 AA.  
 ID 09JEN3

AC 09JEN3:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE RNA POLYMERASE.  
 GN L.  
 OS Tusula paramyxovirus (TPMV).  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Paramyxovirinae; Paramyxovirus.  
 OX NCBI\_TaxID=92129;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Tidona C.A., Darai G.;  
 RT Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF079780; AAF63393.1; -.  
 DR InterPro: IPR001016; -.  
 DR Pfam: PF00946; Paramyx\_RNA\_pol; 1.  
 SQ SEQUENCE 2270 AA; 259395 MW; 037C728466E95D95 CRC64;

Query Match 87.9%; Score 29; DB 14; Length 2270;  
 Best Local Similarity 85.7%; Pred. No. 4.7e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
 |||||  
 DB 351 VNLAEEF 357

RESULT 5  
 O9LDV6 PRELIMINARY: PRT: 245 AA.  
 AC 09LDV6:  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE P0512G09.17 PROTEIN.  
 GN P0512G09.17.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;  
 OC Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone:P0512G09.";  
 RT Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 RT clone:P0695A04.";  
 RT Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AP002836; BAB07935.1; -.  
 DR EMBL: AP002816; BAB03428.1; -.  
 SQ SEQUENCE 245 AA; 26612 MW; 4CD75DC03FA66258 CRC64;

Query Match 84.8%; Score 28; DB 10; Length 245;  
 Best Local Similarity 85.7%; Pred. No. 73;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
 |||||  
 DB 183 VNLAEEF 189

RESULT 6  
 O9HN18



ID 09ANI8 PRELIMINARY: PRT: 280 AA.  
 AC 09ANI8:  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DE PHOSPHONATES TRANSPORT ATP-BINDING.  
 GN PHNC OR VNC2085G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;  
 OC Halobacterium  
 NCBI\_TaxID=64091;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng V.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Strydom J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welt R., Goo Y.A.,  
 RA Leithausen B., Keller K., Cruz R., Danon M.J., Hough D.W.,  
 RA Madocks D.G., Jablonski P.E., Krebs M.P., Angeline C.M., Dale H.,  
 RA Isebarger T.A., Peck R.F., Pohlschroder M., Spidich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassartha S.;  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL; AE005098; AAG20232.1; -  
 DR InterPro: IPR001687; -  
 DR InterPro: IPR002078; -  
 DR InterPro: IPR003439; -  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 KW ATP-binding.  
 SO SEQUENCE 280 AA; 30083 MW; E147D27A4985FA8 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 280;  
 Best Local Similarity 85.7%; Pred. No. 84;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
 DB 199 VNLAEEF 205

RESULT 7  
 P94419 PRELIMINARY: PRT: 315 AA.  
 AC P94419;  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE HOMOLOGUE OF FERREDOXIN ANGIOBACTIN TRANSPORT SYSTEM PERMEASE PROTEIN  
 DE FATC OF V. ANGIOBACTIN.  
 GN YCLO.  
 OS Bacillus subtilis.  
 OC Bacteria: Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 TRPC2;  
 RX MEDLINE=97124189; PubMed=8969502;  
 RA Yamane K., Kumano M., Kurita K.;  
 RT "The 25 degrees-36 degrees region of the Bacillus subtilis chromosome:  
 determination of the sequence of a 146 kb segment and identification  
 of 113 genes";  
 RL Microbiology 142:3047-3056(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertoletti M.G., Bessieres P., Bolotin A., Borchert S.,

RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruns C.V., Caldwell I.B., Capuano V.J., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabre C., Ferrati E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guiseppl G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone P.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viart A., Wambuit R., Wedler E., Wedler H., Weltzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 subtilis";  
 RL Nature 390:249-256(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunze F., Ogasawara N., Yoshikawa H., Danchin A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; D50453; BAA09013.1; -  
 DR EMBL; Z59106; CAB12189.1; -  
 DR InterPro: IPR000522; -  
 DR Pfam: PF01032; FeecD\_family; 1.  
 DR Prodom: PD01557; -; 1.  
 SO SEQUENCE 315 AA; 35473 MW; BCFDEAAE7D92E59 CRC64;

Query Match 84.8%; Score 28; DB 2; Length 315;  
 Best Local Similarity 85.7%; Pred. No. 96;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
 DB 248 VNLAEEF 254

RESULT 8  
 O9PAE1 PRELIMINARY: PRT: 629 AA.  
 ID O9PAE1  
 AC O9PAE1;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE NADPH-DEPENDENT CYTOCHROME P450 OXIDOREDUCTASE (EC 1.6.2.4)  
 OS Cunninghamella echinulata.  
 OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales;  
 OC Cunninghamellaceae; Cunninghamella.  
 OX NCBI\_TaxID=76405;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 26878;  
 RX MEDLINE=20145435; PubMed=10679206;  
 RA Yadav J.S., Loper J.C.;  
 RT "Cloning and characterization of the cytochrome P450 oxidoreductase  
 gene from the zygomycete fungus Cunninghamella";  
 RL Biochem. Biophys. Res. Commun. 268:345-353(2000).  
 DR EMBL; AF195660; AAF89959.1; -

DR InterPro: IPR001094; -  
 DR InterPro: IPR001433; -  
 DR InterPro: IPR001709; -  
 DR InterPro: IPR002028; -  
 DR InterPro: IPR003097; -  
 DR Pfam: PF00175; oxidored\_fad; 1.  
 DR Pfam: PF00667; FAD\_binding; 1.  
 DR PRINTS: PR00369; FLAVODOXIN.  
 DR PRINTS: PR00371; FPNCR.  
 DR PRODOM: PD001535; -; 1.  
 DR Oxidoreductase; -  
 KW NON\_TER 1  
 FT SEQUENCE 629 AA; 71387 MW; EC63589C857965A5 CRC64;

Query Match 84.8%; Score 28; DB 3; Length 629;  
 Best Local Similarity 85.7%; Pred. NO. 2e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEP 7  
 DB 596 VNLAEP 602

RESULT 9  
 O9HEP0 PRELIMINARY; PRT; 888 AA.  
 AC O9HEP0;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE TREHALOSE-6-PHOSPHATE PHOSPHATASE (EC 3.1.3.12).  
 GN TP52.  
 OS Candida albicans (Yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA De Virgilio C.;  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: A1742990; CAC17748.1;  
 KM HYDROLASE. 888 AA; 100407 MW; 157ABDDAA26C64C9 CRC64;  
 SO SEQUENCE

Query Match 84.8%; Score 28; DB 3; Length 888;  
 Best Local Similarity 85.7%; Pred. NO. 3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEP 7  
 DB 634 VNLAEP 640

RESULT 10  
 O17281 PRELIMINARY; PRT; 906 AA.  
 AC O17281;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE CADHERIN HOMOLOG.  
 OS Botryllus schlosseri.  
 CC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
 OC Scolidobranchia; Botryllidae; Botryllus.  
 OX NCBI\_TaxID=30301;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98038982; PubMed=9373145;  
 RA Levi L., Douek J., Osman M., Bosch T.C.G., Rinkovich B.;  
 RA "Cloning and characterization of BS-cadherin, a novel cadherin from  
 RA the colonial urochordate Botryllus schlosseri."

RL Gene 200:117-123(1997).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: TO THE CADHERIN FAMILY.  
 DR EMBL: U61755; AB88396.1;  
 DR HSP. P03803; 1EDH.  
 DR InterPro: IPR00233; -  
 DR InterPro: IPR002126; -  
 DR Pfam: PF00028; cadherin\_4.  
 DR Pfam: PF01049; Cadherin\_C\_term; 1.  
 DR PRINTS: PS00205; CADHERIN.  
 DR PROSITE: PS00232; CADHERIN. 3.  
 DR SMART: SM00112; CA; 1.  
 KW Ca2+ binding; Cell adhesion; Glycoprotein; Repeat; Transmembrane.  
 SO SEQUENCE 906 AA; 100153 MW; B03E3F84F9AF18D2 CRC64;

Query Match 84.8%; Score 28; DB 5; Length 906;  
 Best Local Similarity 85.7%; Pred. NO. 3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEP 7  
 DB 667 VNLAEP 673

RESULT 11  
 O9W215 PRELIMINARY; PRT; 1283 AA.  
 AC O9W215;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE CG10505 PROTEIN.  
 GN CG10505.  
 OS Drosophila melanogaster (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=BERKELEY;  
 RC MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brostein P., Brotler P.,  
 RA Burtis K.C., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA Cherry J.M., Cawley S., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson R., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dunbrin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleschmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasner K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
 RA Jajani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodali C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Port V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spter E., Spredling A.C., Stapleton M., Strong R., Sun E.,

RA Svirsks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:  
 "The genome sequence of *Drosophila melanogaster*."  
 Science 287:2185-2195(2000).  
 RL Science 287:2185-2195(2000).  
 CC -1- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC  
 TRANSPORTERS).  
 CC EMBL: AE003453; AAF46706.1; -.  
 DR HSSP: P13569; 1NBD.  
 DR Flybase: FBgn0034612; CG10505.  
 DR InterPro: IPR001140; -.  
 DR InterPro: IPR001687; -.  
 DR InterPro: IPR003439; -.  
 DR InterPro: IPR003593; -.  
 DR Pfam: PF00605; ABC\_tran. 2.  
 DR Pfam: PF00664; ABC\_membrane. 2.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
 DR SMART: SM00382; AAA; 1.  
 KW ATP-binding; Transport.  
 SO SEQUENCE 1283 AA; 144369 MW; 8864AD3CA54929E CRC64;

Query Match 84.8%; Score 28; DB 5; Length 1283;  
 Best Local Similarity 85.7%; Pred. No. 4.4e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
 |||||  
 DB 1006 VNLAEEW 1012

RESULT 12  
 O9VFC4 PRELIMINARY; PRT; 1332 AA.  
 AC O9VFC4;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE CG6752. PROTEIN.  
 GN CG6752.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriodes; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Gelinkler S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,  
 Wan K.C., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abell J.F., Abmayyan A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 Balcer R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durkin K.J., Evangelista C.C., Ferraz C., Ferrelira S., Fleischmann W.,  
 Fosler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegyan C.,  
 Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Splitter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svirsks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:  
 "The genome sequence of *Drosophila melanogaster*."  
 Science 287:2185-2195(2000).  
 RL Science 287:2185-2195(2000).  
 CC EMBL: AE003707; AAF5136.1; -.  
 DR Flybase: FBgn0038296; CG6752.  
 DR InterPro: IPR00107; -.  
 DR InterPro: IPR001841; -.  
 DR Pfam: PF00622; SPRY; 1.  
 DR SMART: SM00449; SPRY; 1.  
 SO SEQUENCE 1332 AA; 151045 MW; 7DE1B95205BFD60 CRC64;

Query Match 84.8%; Score 28; DB 5; Length 1332;  
 Best Local Similarity 85.7%; Pred. No. 4.6e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
 |||||  
 DB 855 VNAAEEF 861

RESULT 13  
 O9FKH5 PRELIMINARY; PRT; 123 AA.  
 AC O9FKH5;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE SIMILARITY TO TAIL-LIKE NON-LTR RETROELEMENT PROTEIN.  
 OS *Arabidopsis thaliana* (Mouse-ear cress).  
 OC Arabidopsids  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA;  
 RX MEDLINE=98344145; PubMed=9679202;  
 RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,  
 Tabata S.,  
 RT "Structural analysis of *Arabidopsis thaliana* chromosome 5. V. Sequence  
 features of the regions of 1,381,565 bp covered by twenty one  
 RT physically assigned P1 and TAC clones."  
 RT DNA Res. 5:131-145(1998).  
 DR EMBL: AB011485; BAB09272.1; -.  
 SO SEQUENCE 123 AA; 14378 MW; C0C734B59CA9553F CRC64;

Query Match 81.8%; Score 27; DB 10; Length 123;  
 Best Local Similarity 85.7%; Pred. No. 60;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
 |||||  
 DB 111 VNLAEEF 117

RESULT 14  
 085605

ID 085605 PRELIMINARY; PRT; 223 AA.  
 AC 085605;  
 DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE EXCISE/CLASE ABC SUBUNIT A (FRAGMENT).  
 GN UVRA.  
 OS Prevotella albensis.  
 OC Bacteria; CFB group; Bacteroidaceae; Prevotella.  
 OX NCBI\_TaxID-77768;  
 RN [1]  
 RP SEQUENCE OF 9-191 FROM N.A.  
 RC STRAIN-M384;  
 RA Walker N.D., McEwan N.R., Wallace R.J.;  
 RT "Prevotella albensis putative uvra gene."  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS)  
 CC EMBL: AF070990; AAC24131.2; -  
 DR InterPro: IPR003439; -  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
 KW ATP-binding; Transport.  
 FT NON\_TER 1  
 FT NON\_TER 223  
 SQ SEQUENCE 223 AA; 24579 MW; 2563BA0BC2996C9 CRC64;

Query Match 81.8%; Score 27; DB 2; Length 223;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEE 7  
 11:111  
 DB 154 VNMAEF 160

RESULT 15  
 P90781  
 ID P90781 PRELIMINARY; PRT; 251 AA.  
 AC P90781;  
 DT 01-MAY-1997 (TREMblrel. 03, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE C55A6.7 PROTEIN.  
 GN C55A6.7  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID-6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kershaw J.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94150718; PubMed-7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bontfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopia A., Saunders D., Showkeen R.,  
 RA Smailton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Telford M., Thomas K., Vaughan M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Spoat J., Wooldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans."  
 RL Nature 368:32-38(1994).  
 CC -1- SIMILARITY TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)  
 CC FAMILY.  
 CC EMBL: Z81051; CAB02867.1; -  
 DR InterPro: IPR002196; -

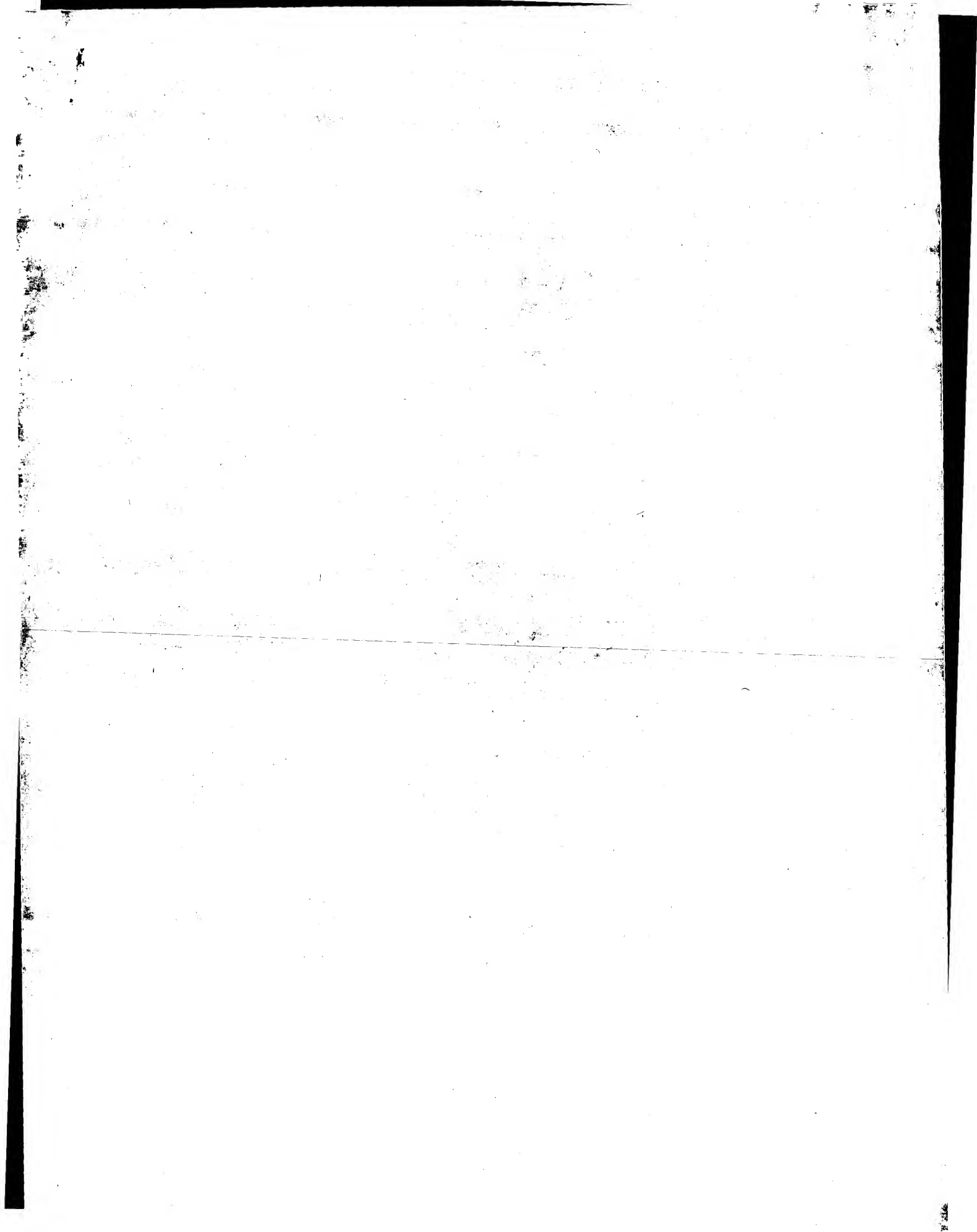
DR InterPro: IPR003015; -  
 DR Pfam: PF00106; adh\_short; 1.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
 KW Oxidoreductase.  
 SQ SEQUENCE 251 AA; 27034 MW; FE1207B860690168 CRC64;

Query Match 81.8%; Score 27; DB 5; Length 251;  
 Best Local Similarity 57.1%; Pred. No. 1.3e+02;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLAEE 7  
 11:111  
 DB 148 INIAEF 154

Search completed: September 6, 2001, 16:49:56  
 Job time: 740 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:10 ; Search time 72.75 Seconds  
(without alignments)  
3.296 Million cell updates/sec

Title: US-09-603-713-27

Sequence: 1 VNLAEF 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	90.9	236	Y385_MYCGE	P47625 mycoplasma
2	28	84.8	289	ATP6_PROMO	P21903 proplionigen
3	28	84.8	1131	YAB9_YEAST	P31380 saccharomyc
4	27	81.8	262	ORC6_MOUSE	O9WUJ8 mus musculu
5	27	81.8	285	AMPR_CITDI	P52658 citrobacter
6	27	81.8	292	AMPR_PROST	O69772 providencia
7	27	81.8	295	NMCR_ENTCL	P52676 enterobacte
8	27	81.8	352	KK4_BRARE	O9PUB8 brachydanio
9	27	81.8	387	SCN1_SCHPO	P41890 schizosacch
10	27	81.8	399	NIFS_KLEPN	P05344 klebsiella
11	27	81.8	449	CSUP_DROME	O9V344 drosophila
12	27	81.8	545	KEAL_CAEEL	O9XRG7 caenorhabdi
13	27	81.8	545	CH60_BORBU	P27575 botrellia du
14	27	81.8	567	MAL1_APIME	Q17058 apis mellif
15	27	81.8	1017	LRG1_YEAST	P35688 saccharomyc
16	27	81.8	1242	RP02_ASF87	P42487 african swi
17	26	78.8	47	VARE_BPP22	P14112 bacteriopho
18	26	78.8	133	AAKG_PIG	O09138 sus scrofa
19	26	78.8	160	RL10_METHH	O27191 methanobact
20	26	78.8	174	RL10_METYA	O57963 methanococ
21	26	78.8	282	ROC_XENLA	P19600 xenopus lae
22	26	78.8	292	SMER_SERMA	P52683 serratia ma
23	26	78.8	303	ROC_HUMAN	P07910 homo sapien
24	26	78.8	313	CBRR_RHIME	P56885 rhizobium m
25	26	78.8	330	AAKG_MOUSE	O54495 mus musculu
26	26	78.8	330	AAKG_MOUSE	P80385 rattus norv
27	26	78.8	331	AAKG_HUMAN	P54619 homo sapien
28	26	78.8	368	TRMU_ECOLI	P25745 escherichia
29	26	78.8	378	URH1_YEAST	O04179 saccharomyc
30	26	78.8	380	COAT_CNV	P15183 cucumber ne
31	26	78.8	399	PNCB_BUCAT	P57442 buchnera ap
32	26	78.8	1065	CARB_BACCL	P46537 bacillus ca
33	25	75.8	165	YG25_HAEIN	P44277 haemophilus

34	25	75.8	170	Y531_METYA	O57951 methanococ
35	25	75.8	244	ANR_PSEAE	P23926 pseudomonas
36	25	75.8	249	GRPE_SYNY3	O59978 synechocyst
37	25	75.8	267	LEU2_ALCEU	O44023 alcaligenes
38	25	75.8	292	TF2_HUMAN	O16635 homo sapien
39	25	75.8	303	LMH1_CHICK	O01635 gallus gall
40	25	75.8	330	OUTG_EMENI	P23416 emericella
41	25	75.8	340	OAX_NEUCR	P11634 neurospora
42	25	75.8	383	TRMU_HAEIN	P44551 haemophilus
43	25	75.8	399	PNCB_ECOLI	P18133 escherichia
44	25	75.8	409	SERA_ECOLI	P08328 escherichia
45	25	75.8	430	MEHF_HAEIN	P44613 haemophilus

## ALIGNMENTS

```

RESULT 1
ID Y385_MYCGE STANDARD: PRT: 236 AA.
AC P47625; Q49354;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MG385.
OS MG385.
OC Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann J.L., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Furumasa J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips G.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bult K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN (2)
RP SEQUENCE OF 3-78 AND 87-176 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bult K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -1- SIMILARITY: TO GLYCEROPHOSPHORYL DIESTER PHOSPHODIESTERASES (EC
CC 3.1.4.46).
CC -1- SIMILARITY: M.GENITALIUM MG293.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U39720; AAC71612.1; -
DR EMBL: U02112; AAC12385.1; -
DR EMBL: U02246; AAA03402.1; -
DR TIGR: MG385; -
KW Hypothetical protein; Hydrolase.
FT CONFLICT 87 87 S -> M (IN REF. 2).
SQ SEQUENCE 236 AA; 27733 MW; D7A3AF4E3630BDE2 CRC64;

```

Query Match 90.9%; Score 30; DB 1; Length 236;  
Best Local Similarity 85.7%; Pred. No. 3.5;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLAEEF 7  
|||||  
DB 75 VNLAEEF 81

## RESULT 2

APR6\_PROMO STANDARD; PRT: 289 AA.  
ID APR6\_PROMO  
AC P21903:  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ATP SYNTHASE A CHAIN, SODIUM ION SPECIFIC (EC 3.6.3.15) (PROTEIN 6).  
GN ATP OR UNCB.  
OS Propionigenium modestum.  
OC Bacteria; Fusobacteria; Propionigenium.  
OX NCBI\_TaxID=2333;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 2376;  
RX MEDLINE=91067471; PubMed=2174545;  
RA Kaim G., Ludwig W., Dimroth P., Schleifer K.H.;  
RT "Sequence of subunits a and b of the sodium ion translocating  
adenosine triphosphate synthase of Propionigenium modestum.";  
RL Nucleic Acids Res. 18:6697-6697(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 2376;  
RX MEDLINE=92339434; PubMed=1386022;  
RA Kaim G., Ludwig W., Dimroth P., Schleifer K.H.;  
RT "Cloning, sequencing and in vivo expression of genes encoding the F0  
part of the sodium-ion-dependent ATP synthase of Propionigenium  
modestum in Escherichia coli.";  
RL Eur. J. Biochem. 207:463-470(1992).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 2376;  
RX MEDLINE=9106937; PubMed=2170948;  
RA Esser U., Krumholz L.R., Simon R.D.;  
RT "Nucleotide sequence of the F0 subunits of the sodium dependent F1F0  
ATPase of Propionigenium modestum.";  
RL Nucleic Acids Res. 18:5887-5888(1990).  
RN [4]  
RP DISCUSSION OF SEQUENCE.  
RX MEDLINE=92249746; PubMed=1533602;  
RA Krumholz L.R., Esser U., Simon R.D.;  
RT "Characterization of the genes coding for the F1F0 subunits of the  
sodium dependent ATPase of Propionigenium modestum.";  
RL FEMS Microbiol. Lett. 70:37-41(1992).  
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL. IT MAY PLAY A  
DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE.  
SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC  
CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE  
SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)  
HAS THREE MAIN SUBUNITS: A, B AND C.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 8  
POTENTIAL TRANSMEMBRANE DOMAINS.  
CC -1- MISCELLANEOUS: THE ATPASE OF P. MODESTUM IS OF SPECIAL INTEREST  
BECAUSE IT USES SODIUM IONS INSTEAD OF PROTONS AS THE  
PHYSIOLOGICAL COUPLING ION.  
CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X54809; CAA38579.1; -.

DR EMBL; X66102; CAA46894.1; -;  
DR EMBL; X53860; CAA37911.1; -;  
DR EMBL; X58461; CAA41368.1; -;  
DR PIR; S12611; S12611.  
DR PIR; S12619; S12619.  
DR PIR; S23334; S23334.  
DR PIR; S36000; S36000.  
DR InterPro; IPR000568; -;  
DR Pfam; PF00119; ATP-synt\_A; 1.  
DR PROSITE; PS00449; ATPASE\_A; 1.  
KW Sodium transport; CF(0); Transmembrane.  
FT CONFLICT 114 114 Y -> I (IN REF. 3).  
SQ SEQUENCE 289 AA; 32224 MM; 6EF6A9BE97A9129E CRC64;

Query Match 84.8%; Score 28; DB 1; Length 289;  
Best Local Similarity 71.4%; Pred. No. 14;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VNLAEEF 7  
|||||  
DB 211 INLAEEF 217

RESULT 3  
YAB9\_YEAST STANDARD; PRT: 1131 AA.  
ID YAB9\_YEAST  
AC P31380:  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE HYPOTHETICAL 128.5 KDA HELICASE IN ATSI-TPD3 INTERGENIC REGION.  
GN YAL019W OR YAL001 OR FUN30.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=9309532; PubMed=8458570;  
RA Ouellette B.F.F., Clark M.W., Keng T., Storms R.K., Zhong W.W.,  
Zeng B., Fortin N., Delaney S., Barton A.B., Kaback D.B., Bussey H.;  
RT "Sequencing of chromosome I from Saccharomyces cerevisiae: analysis  
of a 32 kb region between the LTEL and SP07 genes.";  
RL genome 36:32-42(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=94193531; PubMed=8144453;  
RA Barton A.B., Kaback D.B.;  
RT "Molecular cloning of chromosome I DNA from Saccharomyces cerevisiae:  
analysis of the genes in the FUN38-MAK16-SP07 region.";  
RL J. Bacteriol. 176:1872-1880(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=92221690; PubMed=1561836;  
RA Clark M.W., Zhong W.W., Keng T., Storms R.K., Barton A.B.,  
Kaback D.B., Bussey H.;  
RT "Identification of a Saccharomyces cerevisiae homolog of the SNF2  
transcriptional regulator in the DNA sequence of an 8.6 kb region in  
the LTEL-CYS1 interval on the left arm of chromosome I.";  
RL Yeast 8:133-145(1992).  
CC -1- FUNCTION: MAY FUNCTION AS A NEGATIVE REGULATORY ELEMENT IN THE  
PROCESS OF UV DAMAGE REPAIR.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way



```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L05146; AAC04938.1; -.
DR PIR: S22266; S22266.
DR SGD: S0000017; FUN30.
DR InterPro: IPR000330; -.
DR InterPro: IPR001650; -.
DR Pfam: PF00176; SNE2_N; 1.
DR Pfam: PF00271; Helicase_C; 1.
KW Hypothetical protein; Nuclear protein; DNA-binding; Helicase;
KW ATP-binding.
FT NP_BIND 597 604 ATP (POTENTIAL).
FT SITE 703 706 DECH BOX.
SQ SEQUENCE 1131 AA; 128506 MW; 7E61B9ABB3A42ED2 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 1131;
Best Local Similarity 85.7%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VNLAEE 7
Db 76 VNLAEE 82

RESULT 4
ORC6_MOUSE STANDARD; PRT; 262 AA.
AC 09WUJ8;
DT 01-OCT-2000 (Rel. 40; Created)
DT 01-OCT-2000 (Rel. 40; Last sequence update)
DT 01-OCT-2000 (Rel. 40; Last annotation update)
DE ORIGIN RECOGNITION COMPLEX SUBUNIT 6.
GN ORC6L OR ORC6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Dean F.B., O'Donnell M.;
RT "CDNA Cloning of a homolog for Saccharomyces cerevisiae ORC6 from Mus
RT musculus."
RT Musculus."
RT Submitted (MAR-1999) to the EMBL/Genbank/DDI databases.
CC -1- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT
CC BINDS ORIGIN OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL
CC REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO
CC THE ARS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION IN AN
CC ATP-DEPENDENT MANNER (BY SIMILARITY).
CC -1- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ORC6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF139659; AAD32667.1; -.
CC DNA replication; Nuclear protein; DNA-binding.
CC KW EMBL: AF139659; AAD32667.1; -.
CC SEQUENCE 262 AA; 29188 MW; F8D27EF9C87DE16C CRC64;

Query Match 81.8%; Score 27; DB 1; Length 262;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLAEE 6

```

```

Db .111 VNLAEE 116

RESULT 5
AMPR_CITDI STANDARD; PRT; 285 AA.
ID AMPR_CITDI
AC P52658;
DT 01-OCT-1996 (Rel. 34; Created)
DT 01-OCT-1996 (Rel. 34; Last sequence update)
DT 01-OCT-1996 (Rel. 34; Last annotation update)
DE TRANSCRIPTIONAL ACTIVATOR PROTEIN AMPR.
GN AMPR.
OS Citrobacter diversus.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
OX NCBI_TaxID=545;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ULA27;
RC MEDLINE=94343506; PubMed=8064842;
RX Jones M.E., Avison M.B., Dandinsuren E., Macgowan A.P., Bennett P.M.;
RA "Heterogeneity at the beta-lactamase structural gene ampC amongst
RT Citrobacter spp. assessed by polymerase chain reaction analysis:
RT potential for typing at a molecular level."
RL J. Med. Microbiol. 41:209-214(1994).
CC -1- FUNCTION: THIS PROTEIN IS A POSITIVE REGULATOR OF GENE EXPRESSION
CC OF BETA-LACTAMASE (AMPc).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE LYR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X77656; CA54736.1; -.
CC InterPro: IPR000847; -.
DR Pfam: PF00126; HTH_1; 1.
DR PRINTS: PR00039; HTHLYSR.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Activator.
FT DNA_BIND 22 41 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 285 AA; 32186 MW; 44A670A05057989A CRC64;

Query Match 81.8%; Score 27; DB 1; Length 285;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNLAEE 6
Db 133 VNLAEE 138

RESULT 6
AMPR_PROST STANDARD; PRT; 292 AA.
ID AMPR_PROST
AC 069772;
DT 30-MAY-2000 (Rel. 39; Created)
DT 30-MAY-2000 (Rel. 39; Last sequence update)
DT 30-MAY-2000 (Rel. 39; Last annotation update)
DE TRANSCRIPTIONAL ACTIVATOR PROTEIN AMPR.
GN AMPR.
OS Providencia stuartii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Providencia.
OX NCBI_TaxID=588;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-VDS 96:
RA Koeck J.L., Basmaciogullari S., Parry D., Barnaud G., Teyssou R.,
RA Bulson T., Philippson A., Arlet G.J.,
RT "Cloning and sequencing of ampC and ampR genes from Providencia
RT stuartii."
RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN IS A POSITIVE REGULATOR OF GENE EXPRESSION
CC OF BETA-LACTAMASE (AMPc).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y17315; CAAT6738.1;
DR InterPro: IPR000847;
DR Pfam: PF00126; HTH_1;
DR PRINTS: PR00039; HTHLYSR.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Activator.
FT DNA_BIND 26 45 H-T-H MOTIF (POTENTIAL).
FT SEQUENCE 292 AA; 32741 MW; 9B2CEFF67E112F0 CRC64;
SO
Query Match 81.8%; Score 27; DB 1; Length 292;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VNLAEE 6
DB 137 VNLAEE 142
RESULT 7
NMCR_ENTCL STANDARD; PRT; 295 AA.
ID NMCR_ENTCL
AC P52676;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CARBAPEM-HYDROLYZING BETA-LACTAMASE TRANSCRIPTIONAL ACTIVATOR.
GN NMCR.
OS Enterobacter cloacae.
OC Bacteriia; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxID=550;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NOR-1;
RX MEDLINE=94329582; PubMed=8052644;
RA Naas T., Nordmann P.;
RT "Analysis of a carbapenem-hydrolyzing class A beta-lactamase from
RT Enterobacter cloacae and of its lysr-type regulatory protein."
RT Proc. Natl. Acad. Sci. U.S.A. 91:7693-7697(1994).
CC -1- FUNCTION: THIS PROTEIN IS A POSITIVE REGULATOR OF GENE EXPRESSION
CC OF CARBAPEM-HYDROLYZING BETA-LACTAMASE (NMCA).
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: Z21956; CAAT9966.1;
DR InterPro: IPR000847;
DR Pfam: PF00126; HTH_1; 1.
DR PRINTS: PR00039; HTHLYSR.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Activator.
FT DNA_BIND 22 41 H-T-H MOTIF (POTENTIAL).
FT SEQUENCE 295 AA; 33444 MW; EE3D263AE2BFC7A CRC64;
SO
Query Match 81.8%; Score 27; DB 1; Length 295;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VNLAEE 6
DB 133 VNLAEE 138
RESULT 8
KE4_BRARE STANDARD; PRT; 352 AA.
ID KE4_BRARE
AC O9PUB8;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HISTIDINE-RICH MEMBRANE PROTEIN KE4 HOMOLOG (FRAGMENT).
DE HKE4.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Murray B.W., Snelmann H., Klein J.;
RT "Identification of a homolog of the human HKE4 gene in zebrafish."
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE KE4/CATSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF196345; AAF05821.1;
DR Transmembrane; Glycoprotein.
KW Transmem 3
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 318 338 POTENTIAL.
FT DOMAIN 24 105 HIS-RICH.
FT DOMAIN 177 217 HIS-RICH.
FT CARBOHYD 311 311 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 352 352
FT SEQUENCE 352 AA; 37922 MW; C8C8C60F6D2BA8A6 CRC64;
SO
Query Match 81.8%; Score 27; DB 1; Length 352;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 1 VNLAEE 7
DB 302 VNLAEE 308

```

```

RESULT 9
SCN1_SCHPO STANDARD: PRT: 387 AA.
AC P41890:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SCN1 PROTEIN.
GN SCN1.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95096177; PubMed=7798319;
RA Samejima I., Yanagida M.;
RT "Bypassing anaphase by fission yeast cut9 mutation: requirement of
RT cut9+ to initiate anaphase.";
RL J. Cell Biol. 127:1655-1670(1994).
CC -1- FUNCTION: INTERACTS WITH CUT9.
CC -1- SIMILARITY: TO YEAST YMR262W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D31845; BAA0631.1; -
DR InterPro: IPR001130; -
DR Pfam: PF01026; UPEF0006; 1.
SQ SEQUENCE 387 AA; 44421 MW; AAC154B805BC1085 CRC64;

Query Match 81.8%; Score 27; DB 1; Length 387;
Best Local Similarity 85.7%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7
DB 224 VRLAAEF 230

RESULT 10
NIFS_KLEPN STANDARD: PRT: 397 AA.
AC P05344;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE NIFS PROTEIN.
GN NIFS.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89041575; PubMed=3054814;
RA Beynon J., Cannon M., Banan-Wollaston V., Ally A., Suterquist R.,
RA Cannon F.;
RT "The nucleotide sequence of the nifT, nifY, nifX and nifW genes of K.
RT pneumoniae.";
RL Nucleic Acids Res. 16:9860-9860(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87307977; PubMed=3040672;
RA Beynon J., Ally A., Cannon M., Cannon F., Jacobson M., Cash V.,
RA Dean D.R.;

```

```

RT "Comparative organization of nitrogen fixation-specific genes from
RT Azotobacter vinelandii and Klebsiella pneumoniae: DNA sequence of the
RT nifSV genes.";
RL J. Bacteriol. 169:4024-4029(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89094839; PubMed=3062178;
RA Arnold W., Rump A., Klipp W., Priefer U.B., Puehler A.;
RT "Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the
RT entire nitrogen fixation gene cluster of Klebsiella pneumoniae.";
RL J. Mol. Biol. 203:715-738(1988).
CC -1- FUNCTION: NOT KNOWN YET.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X12600; CAA31118.1; -
DR EMBL: M17350; AAA25156.1; -
DR EMBL: X13303; CAA31675.1; -
DR PIR: S02507; S02507.
DR PIR: S34847; S34847.
DR InterPro: IPR000192; -
DR Pfam: PF00266; AmiNtrn_5; 1.
DR PROSITE: PS00595; AA_TRANSF_5; 1.
KW Nitrogen fixation; Pyridoxal phosphate.
FT BINDING 199
FT CONFLICT 71 79 WPRATPRNA -> CATECHOLAMINES (IN REF. 3).
FT CONFLICT 107 107 L -> M (IN REF. 3).
FT CONFLICT 110 110 Q -> E (IN REF. 3).
FT CONFLICT 121 121 S -> G (IN REF. 3).
FT CONFLICT 191 191 D -> H (IN REF. 2).
FT CONFLICT 202 202 G -> P (IN REF. 2).
FT CONFLICT 287 287 R -> A (IN REF. 3).
SQ SEQUENCE 397 AA; 43254 MW; EE81BB82D0F21F48 CRC64;

Query Match 81.8%; Score 27; DB 1; Length 397;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7
DB 292 VRLAAEF 298

RESULT 11
CSUP_DROME STANDARD: PRT: 449 AA.
AC Q9V3A4;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CATECHOLAMINES UP PROTEIN.
GN CATSUP OR CG10449.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99403013; PubMed=10471719;
RA Stathakis D.G., Burton D.Y., McIvor W.E., Krishnakumar S.,
RA Wright T.R., O'Donnell J.M.;
RT "The catecholamines up (Catsup) protein of Drosophila melanogaster

```

RT Functions as a negative regulator of tyrosine hydroxylase activity."  
 RL Genetics 153:361-382(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MIDDLE-20196006; PubMed-10731132;  
 RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Landell M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Aghayani A., An H.-J., Andrews-Pfanchko C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Goad C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McShenafi A.,  
 RA Merulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,  
 RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan X., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou G., Zhu H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: NEGATIVELY REGULATES TYROSINE HYDROXYLASE ACTIVITY.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE KE4/CAT5UP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF216584; AAB3226.1; -  
 DR EMBL: AE003661; AAF5374.1; -  
 DR Flybase: FBgn0002022; Catsup.  
 KW Transmembrane.  
 FT TRANSMEM 19 39 POTENTIAL.  
 FT TRANSMEM 135 155 POTENTIAL.  
 FT TRANSMEM 167 187 POTENTIAL.  
 FT TRANSMEM 222 242 POTENTIAL.  
 FT TRANSMEM 371 391 POTENTIAL.  
 FT TRANSMEM 395 415 POTENTIAL.  
 FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 449 AA; 48658 MW; F711A234C07AB4C CRC64;

Query Match 81.8%; Score 27; DB 1; Length 449;  
 Best Local Similarity 71.4%; Pred. No. 40;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 UNLAEF 7  
 Db 307 LNLADE 313  
 RESULT 12  
 ID KE4L\_CAEEL STANDARD; PRT; 515 AA.  
 AC 09X107;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE HYPOTHEICAL KE4-LIKE PROTEIN H13N06.5 IN CHROMOSOME X.  
 GN H13N06.5.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 CC Rhabditidae; Poloderinae; Caenorhabditis.  
 CC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE KE4/CAT5UP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: Z99942; CAB17070.1; -  
 DR Wormpep: H13N06.5; CE18815.  
 KW Hypothetical protein; Transmembrane; Glycoprotein.  
 FT TRANSMEM 27 47 POTENTIAL.  
 FT TRANSMEM 214 234 POTENTIAL.  
 FT TRANSMEM 247 267 POTENTIAL.  
 FT TRANSMEM 297 317 POTENTIAL.  
 FT TRANSMEM 386 406 POTENTIAL.  
 FT TRANSMEM 429 449 POTENTIAL.  
 FT TRANSMEM 463 483 POTENTIAL.  
 FT DOMAIN 92 182 HIS-RICH.  
 FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 379 379 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 515 AA; 55500 MW; 17D7E854FE1DAAF CRC64;  
 Query Match 81.8%; Score 27; DB 1; Length 515;  
 Best Local Similarity 71.4%; Pred. No. 46;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 UNLAEF 7  
 Db 370 LNLADE 376  
 RESULT 13  
 ID CH60\_BORBU STANDARD; PRT; 545 AA.  
 AC P27575;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN).  
 GN MPA OR GROEL OR BB0649.  
 OS Borrelia burgdorferi (Lyme disease spirochete).

OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxId-139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91237110; PubMed-1709664;  
 RA Hindersson P., Hansen K.;  
 RT "T cell and antibody reactivity with the Borrelia burgdorferi 60-kDa  
 RL heat shock protein in Lyme arthritis.";  
 RN J. Immunol. 146:3985-3992(1991).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Z57;  
 RX MEDLINE-93014208; PubMed-1356932;  
 RA Wallich R., Helms C., Schable U.E., Lobet Y., Mocer S.E.,  
 RA Kramer M.D., Simon M.M.;  
 RT "Evaluation of genetic divergence among Borrelia burgdorferi isolates  
 RL by use of Ospa, fla, HSP60, and HSP70 gene probes.";  
 RN Infect. Immun. 60:4856-4866(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 35210 / B31;  
 RX MEDLINE-98065943; PubMed-9403685;  
 RA Frazer C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson S., Hanson M.,  
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
 RA Van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
 RA Uitterback T., Matthey L., McDonald L., Artlich P., Bowman C.,  
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
 RA Smith H.O., Venter J.C.;  
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia  
 RL burgdorferi.";  
 RN Nature 390:580-586(1997).  
 CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND  
 CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS  
 CC CONDITIONS (BT SIMILARITY).  
 CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF  
 CC 7 SUBUNITS (BY SIMILARITY).  
 CC -1- DISEASE: THIS PROTEIN IS IMPLICATED IN AUTOIMMUNE DISEASE.  
 CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X54059; CAA37994.1; -  
 CC EMBL: X65139; CAA46269.1; -  
 CC EMBL: AE001166; AAC66995.1; -  
 CC PIR: S19083; S19083.  
 DR HSP: P06139; IGRU.  
 DR TIGR: BB0649; -  
 DR InterPro: IPR001844; -  
 DR Pfam: PF00118; CPN60\_TCP1.1;  
 DR PRINTS: PR00298; CHAPERONIN60.  
 DR PRINTS: PR00304; TCOMPLEXTCP1.  
 DR PROSITE: PS00296; CHAPERONINS\_CPN60.1.  
 DR Chaperone; ATP-binding; Antigen.  
 FT CARBOHYD 124 124 N -> S (IN REF. 1).  
 FT CONFLICT 213 213 N -> S (IN REF. 1).  
 FT CONFLICT 213 213 N -> S (IN REF. 1).  
 SO SEQUENCE 545 AA; 58951 MW; 568BA85BA835719 CRC64;

Query Match 81.8%; Score 27; DB 1; Length 545;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VNLAEE 6  
 |||||

DB 123 VNLAEE 128  
 RESULT 14  
 MAL1\_APIME STANDARD: PRT; 567 AA.  
 ID MAL1\_APIME  
 AC Q17058;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ALPHA-GLUCOSIDASE PRECURSOR (EC 3.2.1.20) (MALTASE).  
 OS Apis mellifera (Honeybee).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Aculeata;  
 OC Apoidea; Apidae; Apis.  
 OX NCBI\_TaxId-7460;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-41; 116-137 AND 255-267.  
 RC TISSUE-hypopharyngeal gland;  
 RX MEDLINE-96192057; PubMed-8619864;  
 RA Ohashi K., Sawata M., Takeuchi H., Natori S., Kubo T.;  
 RT "Molecular cloning of cDNA and analysis of expression of the gene for  
 RL alpha-glucosidase from the hypopharyngeal gland of the honeybee apis  
 RT mellifera L.";  
 RN Biochem. Biophys. Res. Commun. 221:380-385(1996).  
 RP CHARACTERIZATION.  
 RC TISSUE-hypopharyngeal gland;  
 RX MEDLINE-97037073; PubMed-8882720;  
 RA Kubo T., Sasaki M., Nakamura J., Sasagawa H., Ohashi K., Takeuchi H.,  
 RA Natori S.;  
 RT "Change in the expression of hypopharyngeal-gland proteins of the  
 RL worker honeybees (Apis mellifera L.) with age and/or role.";  
 CC J. Biochem. 119:291-295(1996).  
 CC -1- FUNCTION: CONVERTS SUCROSE IN NECTAR TO GLUCOSE AND FRUCTOSE.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING  
 CC 1,4-LINKED D-GLUCOSE RESIDUES WITH RELEASE OF D-GLUCOSE.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE HYPOPHARYNGEAL  
 CC GLANDS OF THE FORAGER (WORKER) HONEYBEE.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO  
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D79208; BAA11466.1; -  
 CC InterPro: IPR000461; -  
 DR Pfam: PF00128; alpha-amylase.1.  
 DR Hydrolase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 18 567  
 FT CARBOHYD 88 88 ALPHA-GLUCOSIDASE.  
 FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 247 247 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 567 AA; 65565 MW; A14D88DD657C99C2 CRC64;

Query Match 81.8%; Score 27; DB 1; Length 567;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VNLAEE 6  
 |||||

Db 447 VNLAAE 452

|||||

RESULT 15

LRG1\_YEAST STANDARD: PRT: 1017 AA.

ID LRG1\_YEAST: 007735;

AC P35688: 01-JUN-1994 (Rel. 29, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE LRG1 PROTEIN.

GN LRG1 OR YDL240W.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

NR [1]

RN SEQUENCE FROM N.A.

RP STRAIN=BY874;

RC MEDLINE=94344779; PubMed=8065929;

RA Mueller A., Xu G., Wells R., Hollenberg C.P., Piepersberg M.;

RT "LRG1 is expressed during sporulation in Saccharomyces cerevisiae and contains motifs similar to LIM and rho/racGAP domains.";

RL Nucleic Acids Res. 22:3151-3154(1994).

RN [2]

RP Alt-Moerle J., Schneider C., Moro M.;

RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: COULD PLAY A ROLE DURING MATING.

CC -1- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED DURING SPORULATION.

CC -1- SIMILARITY: CONTAINS 3 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC IONS.

CC -1- SIMILARITY: CONTAINS 1 GAP DOMAIN.

CC -1- SIMILARITY: SOME, TO GUANINE-NUCLEOTIDE RELEASING FACTORS OF THE CDC24 FAMILY IN THE C-TERMINUS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: X78453; CAA55210.1; -

DR EMBL: Z74288; CAA98820.1; -

DR PIR: S43158; S43158.

DR HSSP: Q07960; IRGP.

DR SGD: S0002399; LRG1.

DR InterPro: IPR000198; -

DR InterPro: IPR001781; -

DR Pfam: PF00620; RhogAP; 1.

DR PROSITE: PS00478; LIM\_DOMAIN\_1; 2.

DR PROSITE: PS50023; LIM\_DOMAIN\_2; 3.

KW Sporulation; LIM motif; Metal-binding; Zinc; Repeat.

KM

FT DOMAIN 28 88 LIM.

FT DOMAIN 98 148 LIM.

FT DOMAIN 135 184 LIM (TRUNCATED).

FT DOMAIN 419 474 LIM.

FT DOMAIN 755 903 GAP DOMAIN.

FT CONFLICT 531 531 H->Q (IN REF. 1).

FT CONFLICT 766 766 R->S (IN REF. 1).

FT CONFLICT 791 791 N->T (IN REF. 1).

FT CONFLICT 821 821 L->Q (IN REF. 1).

FT CONFLICT 838 838 A->S (IN REF. 1).

FT CONFLICT 849 849 V->L (IN REF. 1).

FT CONFLICT 849 849 S->F (IN REF. 1).

FT CONFLICT 895 895 A->T (IN REF. 1).

FT CONFLICT 928 928 Y->S (IN REF. 1).

FT CONFLICT 935 935 INHISTVQSKRTIDYSCDKITPTVTKDSTTVVIGETINK

FT CONFLICT 977 1017 -> TILPPLCKVKQSIIPNVT (IN REF. 1).

SQ SEQUENCE 1017 AA; 116660 MW; DDEB79B4BEBB09 CRC64;

Query Match 81.8%; Score 27; DB 1; Length 1017;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 6; Conservative 0; Mismatches 0; Gaps 0;

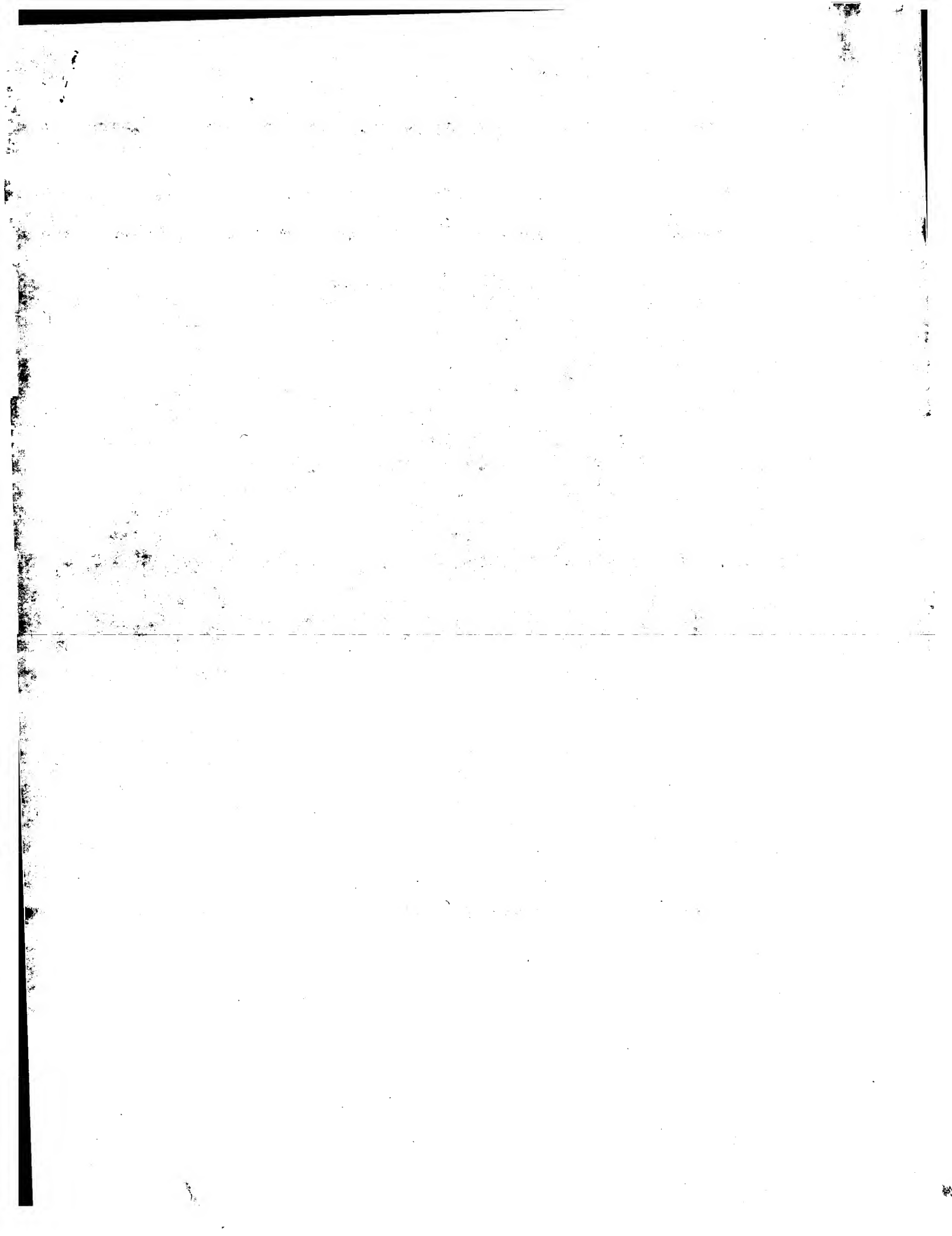
OY 1 VNLAAE 6

DB 193 VNLAAE 198

Search completed: September 6, 2001, 16:51:11

Job time: 815 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:54 ; Search time 134.15 Seconds  
(without alignments)  
3.975 Million cell updates/sec

Title: US-09-603-713-27

Perfect score: 33

Sequence: 1 VNLAEF 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	90.9	236	2 F64242	hypothetical prote
2	29	87.9	328	2 C70368	conserved hypotet
3	28	84.8	280	2 D84358	phosphonates trans
4	28	84.8	289	2 SI2619	Na+-transporting A
5	28	84.8	315	1 C69763	ferrichrome ABC tr
6	28	84.8	1131	2 S22266	FUN30 protein - ye
7	27	81.8	251	2 T20259	hypothetical prote
8	27	81.8	282	2 T03227	hypothetical prote
9	27	81.8	295	2 S35914	regulatory protein
10	27	81.8	387	2 B51164	scn1 protein - its
11	27	81.8	400	2 S02507	nitrogenase cofact
12	27	81.8	500	2 T40666	acetylglutamate sy
13	27	81.8	515	2 T23089	hypothetical prote
14	27	81.8	538	2 E96492	hypothetical prote
15	27	81.8	545	2 H70180	heat shock protein
16	27	81.8	567	2 JCA714	alpha-glucosidase
17	27	81.8	606	1 S59634	endo-1,4-beta-xyla
18	27	81.8	790	1 G69071	heavy-metal-transp
19	27	81.8	1017	2 S67804	LRG1 protein - yea
20	27	81.8	1242	2 S78061	DNA-directed RNA p
21	26	78.8	47	1 W28P22	arf protein - phag
22	26	78.8	160	2 H69015	ribosomal protein
23	26	78.8	174	2 G64367	ribosomal protein
24	26	78.8	225	2 T17795	hypothetical prote
25	26	78.8	241	2 S71473	endo-1,4-beta-xyla
26	26	78.8	282	2 A31765	heterogeneous ribo
27	26	78.8	282	2 T15640	hypothetical prote
28	26	78.8	282	2 A26885	heterogeneous nucl
29	26	78.8	292	2 S49164	Lysr-type protein

30	26	78.8	300	2 G69476	hypothetical prote
31	26	78.8	303	2 C34504	heterogeneous ribo
32	26	78.8	330	2 T10759	AMP-activated prote
33	26	78.8	378	2 S69683	hypothetical prote
34	26	78.8	380	1 VCVGCN	coat protein - cuc
35	26	78.8	383	2 B64858	probable ATPase yc
36	26	78.8	383	2 C85687	hypothetical prote
37	26	78.8	399	2 A84972	nicotinate phospho
38	26	78.8	447	2 E70186	signal recognition
39	26	78.8	451	2 S75569	hypothetical prote
40	26	78.8	466	2 F84632	hypothetical prote
41	26	78.8	473	2 B85187	glycoprotein homol
42	26	78.8	628	2 F81356	methionine-cRNA 1
43	26	78.8	720	2 T51007	hypothetical prote
44	26	78.8	881	2 D83398	probable two-compo
45	26	78.8	884	2 H83322	hypothetical prote

## ALIGNMENTS

```

RESULT 1
F64242
hypothetical protein MG385 - Mycoplasma genitalium
C:Species: Mycoplasma genitalium
C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000
C:Accession: F64242
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.
M.; Fuhrmann, J.; Nguyen, D.; Uitterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346
A:Accession: F64242
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-236 <TIGR>
A:Cross-references: GB:U39723; GB:I43967; NID:g1046092; PID:g1046096; TIGR:MG385
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: glycerophosphodiester phosphodiesterase

Query Match          90.9%; Score 30; DB 2; Length 236;
Best Local Similarity 85.7%; Pred No. 8.9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 VNLAEF 7
      |||||
DB      75 VNLAEF 81

RESULT 2
C70368
conserved hypothetical protein aq_778 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Sep-1999
C:Accession: C70368
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: C70368
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-328 <AOE>
A:Cross-references: GB:AE000707; NID:g2983342; PIDN:AAC06942.1; PID:g2983354; GB:AE000
C:Genetics:
A:Gene: aq_778
C:Superfamily: Escherichia coli ybhk protein

```

Query Match 87.9%; Score 29; DB 2; Length 328;  
 Best Local Similarity 85.7%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
 DB 133 VNLAEEF 139

RESULT 3  
 D84358  
 Phosphonates transport ATP-binding [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: D84358  
 R:Ng, W.V.; Kennedy, S.P.; Mahatras, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leitauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jadic, Jung, K.H.; Alam, M.; Freitas, T. 12176-12181, 2000  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483  
 A:Accession: D84358  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-280 <STO>  
 A:Cross-references: GB:AE004437; NID:g10581501; PIDN:AG20323.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: phnc

Query Match 84.8%; Score 28; DB 2; Length 280;  
 Best Local Similarity 85.7%; Pred. No. 33;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
 DB 199 VNLAEEF 205

RESULT 4  
 S12619  
 Na<sup>+</sup>-transporting ATP synthase (F<sub>0</sub> 3.6.1.-) chain a - *Propionigenium modestum*  
 C:Species: *Propionigenium modestum*  
 C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 20-Jun-2000  
 C:Accession: S12619; S29035; S23334; S24367; S36000  
 R:Kaim, G.; Ludwig, W.; Dimroth, P.; Schleifer, K.H.  
 Nucleic Acids Res. 18, 6697, 1990  
 A:Title: Sequence of subunits a and b of the sodium ion translocating adenosine triphosph  
 A:Reference number: S12619; MUID:91067471  
 A:Accession: S12619  
 A:Molecule type: DNA  
 A:Residues: 1-289 <KAI>  
 A:Cross-references: EMBL:X54809; NID:g45607; PIDN:CAA38579.1; PID:g45608  
 R:Esser, U.; Krumholz, L.R.; Simoni, R.D.  
 Nucleic Acids Res. 18, 5887, 1990  
 A:Title: Nucleotide sequence of the F<sub>0</sub>(O) subunits of the sodium dependent F<sub>1</sub>F<sub>0</sub> ATPase  
 A:Reference number: S12611; MUID:91016937  
 A:Accession: S12611  
 A:Status: translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-113, 'I', 115-289 <ESS>  
 A:Cross-references: EMBL:X53960; NID:g45643; PIDN:CAA37911.1; PID:g45644  
 R:Krumholz, L.R.; Esser, U.; Simoni, R.D.  
 FEBS Microbiol. Lett. 91, 37-42, 1992  
 A:Title: Characterization of the genes coding for the F<sub>1</sub>F<sub>0</sub>(O) subunits of the sodium de  
 A:Reference number: S29034  
 A:Accession: S29035  
 A:Molecule type: DNA  
 A:Residues: 1-113, 'I', 115-287 <KRUI>  
 A:Cross-references: EMBL:X58461

R:Krumholz, L.R.  
 submitted to the EMBL Data Library, March 1991  
 A:Reference number: S23334  
 A:Accession: S23334  
 A:Molecule type: DNA  
 A:Residues: 1-113, 'I', 115-289 <KRUI>  
 A:Cross-references: EMBL:X58461; NID:g897797; PIDN:CAA41368.1; PID:g45649  
 R:Kaim, G.; Ludwig, W.; Dimroth, P.; Schleifer, K.H.  
 Eur. J. Biochem. 207, 463-470, 1992  
 A:Title: Cloning, sequencing and in vivo expression of genes encoding the F<sub>0</sub>(O) part o  
 A:Reference number: S24366; MUID:92339434  
 A:Accession: S24367  
 A:Molecule type: DNA  
 A:Residues: 1-95, 'L', 97-157, 'T', 158-205, 'L', 207-224, 'F', 226-289 <KAW>  
 A:Cross-references: EMBL:X66102  
 R:Ludwig, W.  
 submitted to the EMBL Data Library, May 1992  
 A:Reference number: S36000  
 A:Accession: S36000  
 A:Molecule type: DNA  
 A:Residues: 1-289 <LUD>  
 A:Cross-references: EMBL:X66102; NID:g45599; PIDN:CAA46894.1; PID:g45601  
 C:Genetics:  
 A:Gene: uncB  
 A:Superfamily: H<sup>+</sup>-transporting ATP synthase protein 6  
 C:Keywords: ATP biosynthesis; blocked amino end; hydrolase; membrane-associated comp

Query Match 84.8%; Score 28; DB 2; Length 289;  
 Best Local Similarity 71.4%; Pred. No. 35;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
 DB 211 INLAEEF 217

RESULT 5  
 C69763  
 ferrichrome ABC transporter (permease) homolog yciO - *Bacillus subtilis*  
 C:Species: *Bacillus subtilis*  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
 C:Accession: C69763  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
 C:Bron, S.; Brouillet, S.; Busch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
 A:Erlich, S.D.; Emerson, P.T.; Estlin, K.D.; Errington, J.; Fabret, C.; Ferrari,  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
 lech, J.; Harwood, C.R.; Hentzel, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M  
 Koeltter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Mau  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma  
 Y., M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet  
 Rieger, M.; Rivolta, C.; Roche, B.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl  
 A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
 A:Kuchl, M.; Tamakoshi, A.; Tanaka, T.; Terpiltra, P.; Toiguchi, A.; Tosato, V.; Uchiya  
 T.; Walters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
 A:Authors: Yoshikawa, H.F.; Zumbato, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: C69763  
 A:Molecule type: DNA  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Residues: 1-315 <RUN>  
 A:Cross-references: GB:299106; GB:AL009126; NID:g2632653; PIDN:CAB12189.1; PID:g26326  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yciO  
 C:Superfamily: ferrichrome ABC transporter

Query Match 84.8%; Score 28; DB 1; Length 315;  
 Best Local Similarity 85.7%; Pred. No. 38;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEP 7  
1111 11  
DB 248 VNLAEP 254

## RESULT 6

S22266  
FUNG3 protein - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YAL001; protein YAL019w  
C:Species: Saccharomyces cerevisiae  
C>Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 29-Oct-1999  
C:Accession: S22266; S36715  
R:Clark, M.W.; Zhong, W.W.; Keng, T.; Storms, R.K.; Barton, A.; Kaback, D.B.; Bussey, H.  
Yeast 8, 133-145, 1992  
A:Title: Identification of a Saccharomyces cerevisiae homolog of the SNF2 transcription factor  
A:Reference number: S22266; MUID:92221690  
A:Accession: S22266  
A:Molecule type: DNA  
A:Residues: 1-1131 <CLL>  
A:Cross-references: GB:L05146; EMBL:S93805; NID:g171851; PIDN:AAC04938.1; PID:g171856  
R:Ouellette, F.; Clark, M.W.; Keng, T.; Storms, R.K.; Zhong, W.; Zeng, B.; Fortin, N.; D  
submitted to the EMBL Data Library, January 1993  
A:Description: Sequencing of Chromosome I from Saccharomyces cerevisiae: analysis of a 3  
A:Reference number: S36711  
A:Accession: S36715  
A:Molecule type: DNA  
A:Residues: 1-1131 <OU>  
A:Cross-references: EMBL:L05146; NID:g171851; PIDN:AAC04938.1; PID:g171856; MIPS:YAL019w  
C:Genetics:  
A:Gene: SGD:FUN30  
A:Cross-references: SGD:S000017; MIPS:YAL019w  
A:Map position: 1L  
C:Keywords: DNA binding; nucleus; transmembrane protein  
F:969-985/Domain: transmembrane #status predicted <TML>

Query Match 84.8%; Score 28; DB 2; Length 1131;  
Best local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEP 7  
1111 11  
DB 76 VNLAEP 82

## RESULT 7

T20259  
hypothetical protein C55A6.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
C:Accession: T20259  
R:Kershaw, J.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19243  
A:Accession: T20259  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-251 <HIL>  
A:Cross-references: EMBL:Z81051; PIDN:CAB02867.1; GSPDB:GN00023; CESP:C55A6.7  
A:Experimental source: clone C55A6  
C:Genetics:  
A:Gene: CESP:C55A6.7  
A:Map position: 5  
A:Introns: 43/3; 218/3  
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 81.8%; Score 27; DB 2; Length 251;  
Best local Similarity 57.1%; Pred. No. 53;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLAEP 7

DB 148 INTASEP 154  
111111

## RESULT 8

T03227  
hypothetical protein - Streptomyces hygroscopicus  
C:Species: Streptomyces hygroscopicus  
C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 22-Oct-1999  
C:Accession: T03227  
R:Ruan, X.; Stassi, D.; Lax, S.; Katz, L.  
Gene 203, 1-9, 1997  
A:Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus AT  
A:Reference number: Z14848; MUID:98085969  
A:Accession: T03227  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-282 <KUA>  
A:Cross-references: EMBL:AF007101; NID:g2624946; PIDN:AAC38067.1; PID:g2624954  
A:Experimental source: ATCC 29253

Query Match 81.8%; Score 27; DB 2; Length 282;  
Best local Similarity 100.0%; Pred. No. 60;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLAEP 6  
111111  
DB 112 VNLAEP 117

## RESULT 9

S35914  
regulatory protein NOR-R - Enterobacter cloacae  
C:Species: Enterobacter cloacae  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 29-Sep-1999  
C:Accession: S35914  
R:Naas, T.; Nordmann, P.; Mariotte, S.; Arber, W.  
submitted to the EMBL Data Library, March 1993  
A:Description: A carboxypeptidase-hydrolysing beta-lactamase from Enterobacter cloacae NOR-  
A:Reference number: S35914  
A:Accession: S35914  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-295 <NAA>  
A:Cross-references: EMBL:Z21956; NID:g396086; PIDN:CAA79966.1; PID:g396087  
C:Superfamily: regulatory protein ampr

Query Match 81.8%; Score 27; DB 2; Length 295;  
Best local Similarity 100.0%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLAEP 6  
111111  
DB 133 VNLAEP 138

## RESULT 10

B55164  
scn1 protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 23-Mar-1995 #sequence\_revision 23-Mar-1995 #text\_change 29-Oct-1999  
C:Accession: B55164  
R:Samejima, T.; Yanagida, M.  
J. Cell Biol. 127, 1655-1670, 1994  
A:Title: Bypassing anaphase by fission yeast cut9 mutation: requirement of cut9(+) to  
A:Reference number: A55164; MUID:95096177  
A:Accession: B55164  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-387 <SAM>  
A:Cross-references: GB:D31845; NID:g633128; PIDN:BA06631.1; PID:d1007203; PID:g63312

C:Genetics:  
A:Gene: scm1

Query Match 81.8%; Score 27; DB 2; Length 387;  
Best Local Similarity 85.7%; Pred. No. 84;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
| | | | |  
DB 224 VNLAEEF 230

## RESULT 11

S02507 nitrogenase cofactor synthesis protein nifs - Klebsiella pneumoniae  
M:Alternate names: nifs protein  
M:Contains: L-cysteine sulfintransferase (EC 2.8.1.-)  
C:Species: Klebsiella pneumoniae  
C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 23-Mar-2001  
C:Accession: S02507; S01708; S34847; S37296  
R:Arnold, W.; Rump, A.; Klipp, W.; Pfeifer, U.B.; Puchler, A.  
J. Mol. Biol. 203, 715-738, 1988  
A:Title: Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nifH  
A:Reference number: S01836; MUID:89094839  
A:Accession: S02507  
A:Molecule type: DNA  
A:Residues: 1-400 <ARN>  
A:Cross-references: EMBL:X13303; NID:943820; PIDN:CA31675.1; PID:943831  
R:Baynon, J.; Cannon, M.; Buchanan-Wollaston, V.; Alty, A.; Setterquist, R.; Dean, D.; C  
Nucleic Acids Res. 16, 9860, 1988  
A:Title: The nucleotide sequence of the nifH, nifV, nifX and nifW genes of K. pneumoniae  
A:Reference number: S01702; MUID:89041575  
A:Accession: S01708  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-70, 'WPRATPRHA', 83-109, 'L', 111-112, 'Q', 114-123, 'S', 125-289, 'R', 291-400 <BEY  
A:Cross-references: EMBL:X12600; NID:943877; PIDN:CA31118.1; PID:943880  
A:Note: this sequence was submitted to the EMBL Data Library, Aug-1988  
R:Baynon, J.; Alty, A.; Cannon, M.; Cannon, F.; Jacobson, M.; Cash, V.; Dean, D.  
J. Bacteriol. 169, 4024-4029, 1987  
A:Title: Comparative organization of nitrogen fixation-specific genes from Azotobacter  
A:Reference number: S29756; MUID:87307977  
A:Accession: S34847  
A:Molecule type: DNA  
A:Residues: 1-70, 'WPR', 77-78, 'PRHA', 83-109, 'L', 111-112, 'Q', 114-123, 'S', 125-289, 'R', 291-4  
A:Cross-references: EMBL:M17350  
R:Baynon, J.; Alty, A.; Cannon, M.; Jacobson, F.; Jacobson, M.; Cash, V.; Dean, D.  
submitted to the EMBL Data Library, October 1987  
A:Description: Comparative organization of nitrogen fixation-specific genes from Azotoba  
A:Reference number: S34843  
A:Accession: S37296  
A:Molecule type: DNA  
A:Residues: 1-70, 'WPR', 77-78, 'PRHA', 83-109, 'L', 111-112, 'Q', 114-123, 'S', 125-193, 'H', 195-2  
A:Cross-references: EMBL:M17350; NID:9149342; PIDN:AAA2156.1; PID:9149344  
C:Genetics:  
A:Gene: nifs  
C:Superfamily: nitrogen fixation protein nifs  
C:Keywords: phosphoprotein; pyridoxal phosphate; sulfuryltransferase  
F:202/Binding site: pyridoxal phosphate (Lys) (covalently #status predicted  
F:324/Active site: Cys (cysteine persulfide Intermediate) #status predicted

## Query Match

81.8%; Score 27; DB 2; Length 400;  
Best Local Similarity 85.7%; Pred. No. 86;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
| | | | |  
DB 295 VNLAEEF 301

## RESULT 12

T40666 acetylglutamate synthase - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T40666  
R:Lyme, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z21889  
A:Accession: T40666  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-500 <LXN>  
A:Cross-references: EMBL:AL034352; PIDN:CAA22186.1; GSPDB:GN00067; SPDB:SPBC725.14  
C:Genetics:  
A:Gene: SPDB:SPBC725.14  
A:Map position: 2

## Query Match

81.8%; Score 27; DB 2; Length 500;  
Best Local Similarity 71.4%; Pred. No. 11e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
| | | | |  
DB 242 INLAOEF 248

## RESULT 13

T23089 hypothetical protein H13N06.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T23089  
R:Lennard, N.  
submitted to the EMBL Data Library, October 1997  
A:Reference number: Z19673  
A:Accession: T23089  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-515 <MLN>  
A:Cross-references: EMBL:Z99942; PIDN:CAH17070.1; GSPDB:GN00028; CESP:H13N06.5  
A:Experimental source: clone H13N06  
C:Genetics:  
A:Gene: CESP:H13N06.5  
A:Map position: X  
A:Insertions: 118/1; 156/2; 182/2; 306/1

## Query Match

81.8%; Score 27; DB 2; Length 515;  
Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
| | | | |  
DB 370 INLAADF 376

## RESULT 14

E96492 hypothetical protein FSA13.3 [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: E96492  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marcia  
Rizzo, M.; Rooney, F.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: E96492  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-538 <ST0>  
 A:Cross-references: GB:AE005173; NID:g9802762; PIDN:AAF59831.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F5A13.3  
 A:Map position: 1

Query Match 81.8%; Score 27; DB 2; Length 538;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLA6 6  
 DB 533 VNLA6 538

RESULT 15

H70180  
 heat shock protein HSP60 - Lyme disease spirochete  
 N:Alternate names: heat shock protein groEL  
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
 C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 23-Mar-2001  
 C:Accession: H70180; S16428; S19083; A49209; S20619  
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
 son, D.; Peterson, J.; Kierliage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
 Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A>Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
 A:Reference number: A70100; MUID:98065943  
 A:Accession: H70180  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-545 <K15>  
 A:Cross-references: GB:AE001166; GB:AE000783; NID:g268857.; PIDN:AAC66995.1; PID:g268857  
 A:Experimental source: strain B31  
 R:Shanfield, M.C.; Hindersson, P.; Soderberg, C.; Mensl, N.; Turck, C.W.; Webb, D.; Ysse  
 J. Immunol. 146, 3985-3992, 1991  
 A>Title: T cell and antibody reactivity with the Borrelia burgdorferi 60-kDa heat shock  
 A:Reference number: S16428; MUID:91237110  
 A:Accession: S16428  
 A:Molecule type: DNA  
 A:Residues: 1-123,'S',125-199,'T',201-212,'S',214-545 <SHA>  
 A:Cross-references: EMBL:X54059; NID:g39364; PIDN:CAA37994.1; PID:g39365  
 A>Note: the authors translated the codon ACT for residue 200 as Ser  
 R:Hindersson, P.  
 A:Description: Nucleotide sequence of the Treponema pallidum analogue to the 65 kDa anti  
 A:Reference number: S19083  
 A:Accession: S19083  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-123,'S',125-199,'X',201-212,'S',214-545 <HIN>  
 A:Cross-references: EMBL:X54059; NID:g39364; PIDN:CAA37994.1; PID:g39365  
 R:Wallich, R.; Helmes, C.; Schallbe, U.E.; Lobet, Y.; Moter, S.E.; Kramer, M.D.; Simon,  
 Infect. Immun. 60, 4856-4866, 1992  
 A>Title: Evaluation of genetic divergence among Borrelia burgdorferi isolates by use of  
 A:Reference number: A49209; MUID:93014208  
 A:Accession: A49209  
 A:Molecule type: DNA  
 A:Residues: 1-545 <MA1>  
 A:Cross-references: EMBL:X65139; NID:g39373; PIDN:CAA46269.1; PID:g39374  
 A:Experimental source: strain Z57  
 A>Note: sequence extracted from NCBI backbone (NCBIP:116498)  
 C:Superfamily: chaperonin groEL

Query Match 81.8%; Score 27; DB 2; Length 545;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VNLA6 6  
 DB 123 VNLA6 128

Search completed: September 6, 2001, 16:45:54  
 Job time: 498 sec

Fri Sep 7 10:58:08 2001

us-09-603-713-27.rpt

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:39 : Search time 113.12 Seconds  
(without alignments)  
1.274 Million cell updates/sec

Title: US-09-603-713-27

Perfect score: 33

Sequence: 1 VNLAEF 7

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/CTUS.COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	87.9	657	4 US-09-306-593-2	Sequence 2, Appl
2	27	81.8	9	3 US-08-802-981-219	Sequence 219, App
3	27	81.8	10	2 US-08-659-984A-19	Sequence 19, Appl
4	27	81.8	10	4 US-08-660-531-19	Sequence 19, Appl
5	27	81.8	11	5 PCT-US94-07043A-3	Sequence 3, Appl
6	27	81.8	21	3 US-08-659-984A-18	Sequence 18, Appl
7	27	81.8	21	3 US-08-802-981-112	Sequence 112, App
8	27	81.8	21	4 US-08-660-531-18	Sequence 18, Appl
9	27	81.8	30	4 US-08-659-984A-17	Sequence 17, Appl
10	27	81.8	30	4 US-08-660-531-17	Sequence 17, Appl
11	27	81.8	33	4 US-08-659-984A-16	Sequence 16, Appl
12	27	81.8	33	4 US-08-660-531-16	Sequence 16, Appl
13	27	81.8	42	2 US-08-659-984A-15	Sequence 15, Appl
14	27	81.8	42	4 US-08-660-531-15	Sequence 15, Appl
15	27	81.8	47	1 US-08-415-751-12	Sequence 12, Appl
16	27	81.8	103	3 US-08-339-708A-12	Sequence 12, Appl
17	27	81.8	493	3 US-09-090-808A-4	Sequence 4, Appl
18	27	81.8	506	2 US-08-659-984A-21	Sequence 21, Appl
19	27	81.8	506	4 US-08-660-531-21	Sequence 21, Appl
20	26	78.8	17	3 US-09-101-146-52	Sequence 52, Appl
21	26	78.8	328	4 US-08-878-989-7	Sequence 7, Appl
22	26	78.8	328	4 US-09-272-796-7	Sequence 7, Appl
23	26	78.8	331	2 US-08-878-989-21	Sequence 21, Appl
24	26	78.8	331	3 US-09-101-146-64	Sequence 64, Appl
25	26	78.8	331	4 US-09-272-796-21	Sequence 21, Appl
26	25	75.8	37	1 US-08-244-491A-8	Sequence 8, Appl
27	25	75.8	37	2 US-08-797-727A-8	Sequence 8, Appl

28	25	75.8	40	4 US-07-861-458C-118	Sequence 118, App
29	25	75.8	45	1 US-08-244-491A-10	Sequence 10, Appl
30	25	75.8	45	2 US-08-797-727A-10	Sequence 10, Appl
31	25	75.8	45	1 US-08-797-727A-20	Sequence 20, Appl
32	25	75.8	46	1 US-08-244-491A-12	Sequence 12, Appl
33	25	75.8	46	2 US-08-797-727A-12	Sequence 12, Appl
34	25	75.8	46	2 US-08-797-727A-19	Sequence 19, Appl
35	25	75.8	47	1 US-08-244-491A-3	Sequence 3, Appl
36	25	75.8	47	2 US-08-797-727A-3	Sequence 3, Appl
37	25	75.8	47	2 US-08-797-727A-18	Sequence 18, Appl
38	25	75.8	47	4 US-08-411-760-5	Sequence 5, Appl
39	25	75.8	48	2 US-08-797-727A-17	Sequence 17, Appl
40	25	75.8	49	2 US-08-797-727A-16	Sequence 16, Appl
41	25	75.8	50	2 US-08-797-727A-15	Sequence 15, Appl
42	25	75.8	51	2 US-08-797-727A-14	Sequence 14, Appl
43	25	75.8	51	4 US-08-411-760-2	Sequence 2, Appl
44	25	75.8	52	1 US-08-244-491A-2	Sequence 2, Appl
45	25	75.8	52	2 US-08-797-727A-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-09-306-593-2  
; Sequence 2, Application US/09306593  
; Patent No. 6184018  
; GENERAL INFORMATION:  
; APPLICANT: Li, Xin-Liang  
; APPLICANT: Ljungdahl, Lars G.  
; APPLICANT: Chen, Huizhong  
; APPLICANT: Ximenes, Eduardo A.  
; TITLE OF INVENTION: Beta-glucosidase Coding Sequences and Protein from  
; FILE REFERENCE: 31-98us  
; CURRENT APPLICATION NUMBER: US/09/306,593  
; CURRENT FILING DATE: 1999-05-06  
; EARLIER APPLICATION NUMBER: US 60/084,494  
; EARLIER FILING DATE: 1998-05-06  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 657  
; TYPE: PRT  
; ORGANISM: Orpinomyces sp. PC-2  
US-09-306-593-2

Query Match 87.9%; Score 29; DB 4; Length 657;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 NLAAEF 7  
| | | | |  
Db 377 NLAAEF 382

RESULT 2  
US-08-802-981-219  
; Sequence 219, Application US/08802981  
; Patent No. 6037137  
; GENERAL INFORMATION:  
; APPLICANT: Komoriya, Akira  
; APPLICANT: Packard, Beverly S.  
; TITLE OF INVENTION: Compositions for the Detection of Enzyme  
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 231  
; CORRESPONDENCE ADDRESSES:  
; ADDRESS: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA

ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-000300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 219:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-802-981-219

Query Match 81.8%; Score 27; DB 3; Length 9;  
Best Local Similarity 85.7%; Pred. No. 1.4e+05;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VNLAEEF 7  
||| |||  
Db 3 VNLAEEF 9

RESULT 3  
US-08-659-984A-19  
Sequence 19, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422

INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Region  
LOCATION: one-of(1)  
OTHER INFORMATION: /note= "N-terminal Ser is acetylated."  
US-08-659-984A-19

Query Match 81.8%; Score 27; DB 2; Length 10;  
Best Local Similarity 85.7%; Pred. No. 0.9; 1; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VNLAEEF 7  
||| |||  
Db 3 VNLAEEF 9

RESULT 4  
US-08-660-531-19  
Sequence 19, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Kelm, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Region  
LOCATION: one-of(1)  
OTHER INFORMATION: /note= "N-terminal Ser is acetylated."  
US-08-660-531-19



Query Match 81.8%; Score 27; DB 4; Length 10;  
Best Local Similarity 85.7%; Pred. No. 0.9;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
111111  
DB 3 VNLAEEF 9

RESULT 5  
PCT-US94-07043A-3  
Sequence 3, Application PC/TUS9407043A

GENERAL INFORMATION:  
APPLICANT: Tamburini, Paul P.; Benz, G nter; H bich,  
APPLICANT: Dieter; Dreyer, Robert N.; Koenig, Gerhard  
TITLE OF INVENTION: CATHEPSIN D IS AN AMYLOIDGENIC  
TITLE OF INVENTION: PROTEASE IN ALZHEIMER S DISEASE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Miles Inc.  
STREET: 400 Morgan Lane  
CITY: West Haven  
STATE: Connecticut  
COUNTRY: USA  
ZIP: 06516

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
COMPUTER: Sharp PC 4600  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/07043A  
FILING DATE: June 21, 1994

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/10889  
FILING DATE: November 12, 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/995,660  
FILING DATE: December 16, 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/880,914  
FILING DATE: May 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pamela A. Simonton  
REGISTRATION NUMBER: 31,060  
REFERENCE/DOCKET NUMBER: MTI 224.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (203) 937-2340  
TELEFAX: (203) 937-2795  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
PCT-US94-07043A-3

Query Match 81.8%; Score 27; DB 5; Length 11;  
Best Local Similarity 85.7%; Pred. No. 1;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
111111  
DB 4 VNLAEEF 10

RESULT 6  
US-08-659-984A-18  
Sequence 18, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:

APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Croak, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-00281005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-18

Query Match 81.8%; Score 27; DB 2; Length 21;  
Best Local Similarity 85.7%; Pred. No. 2.1;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
111111  
DB 3 VNLAEEF 9

RESULT 7  
US-08-802-981-112  
Sequence 112, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:

APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly S.  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-00030005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 112:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product= "Aib"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /product= "Acp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 16  
OTHER INFORMATION: /product= "Acp"  
US-08-802-981-112

Query Match 81.8%; Score 27; DB 3; Length 21;  
Best Local Similarity 85.7%; Pred. No. 2.1;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VNLAEEF 7  
||| |||  
DB 8 VNLDAAEF 14

RESULT 8  
US-08-660-531-18  
Sequence 18, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:

NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-00221005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-18

Query Match 81.8%; Score 27; DB 4; Length 21;  
Best Local Similarity 85.7%; Pred. No. 2.1;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VNLAEEF 7  
||| |||  
DB 3 VNLDAAEF 9

RESULT 9  
US-08-659-984A-17  
Sequence 17, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Jacobson-Croak, Kirsten L.  
APPLICANT: Sinha, Sukanto  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-00281005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-659-984A-17

Query Match 81.8%; Score 27; DB 2; Length 30;

Best Local Similarity 85.7%; Pred. No. 3.2;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAAEF 7  
DB 24 VNLDAEF 30

RESULT 10  
US-08-660-531-17

; Sequence 17, Application US/08660531  
; Patent No. 6221645  
; GENERAL INFORMATION:  
; APPLICANT: Chrysler, Susanna M.S.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Keim, Pamela S.  
; APPLICANT: Anderson, John P.  
; TITLE OF INVENTION: Beta-Secretase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/660,531  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/480,498  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 15270-002100S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-660-531-17

Query Match 81.8%; Score 27; DB 4; Length 30;  
Best Local Similarity 85.7%; Pred. No. 3.2;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAAEF 7  
DB 24 VNLDAEF 30

RESULT 11  
US-08-659-984A-16

; Sequence 16, Application US/0865984A  
; Patent No. 5942400  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, John P.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Jacobson-Croak, Kirsten L.  
; TITLE OF INVENTION: Assays for Detecting Beta-Secretase

; TITLE OF INVENTION: Inhibition  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/659,984A  
; FILING DATE: 07-JUN-1996  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,152  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 15270-0028100S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 33 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-659-984A-16

Query Match 81.8%; Score 27; DB 2; Length 33;  
Best Local Similarity 85.7%; Pred. No. 3.6;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAAEF 7  
DB 15 VNLDAEF 21

RESULT 12

; US-08-660-531-16  
; Sequence 16, Application US/08660531  
; Patent No. 6221645  
; GENERAL INFORMATION:  
; APPLICANT: Chrysler, Susanna M.S.  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Keim, Pamela S.  
; APPLICANT: Anderson, John P.  
; TITLE OF INVENTION: Beta-Secretase  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Ctr., 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/660,531  
; FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-16

Query Match 81.8%; Score 27; DB 4; Length 33;  
Best Local Similarity 85.7%; Pred. No. 3.6;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEE 7  
||| |||  
Db 15 VNLAEE 21

RESULT 13  
US-08-659-984A-15  
Sequence 15, Application US/08659984A  
Patent No. 5942400  
GENERAL INFORMATION:  
APPLICANT: Anderson, John P.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Jacobson-Crook, Kirsten L.  
TITLE OF INVENTION: Assays for Detecting Beta-Secretase  
TITLE OF INVENTION: Inhibition  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/659,984A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,152  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: linear

MOLECULE TYPE: peptide  
US-08-659-984A-15

Query Match 81.8%; Score 27; DB 2; Length 42;  
Best Local Similarity 85.7%; Pred. No. 4.7;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEE 7  
||| |||  
Db 24 VNLAEE 30

RESULT 14  
US-08-660-531-15  
Sequence 15, Application US/08660531  
Patent No. 6221645  
GENERAL INFORMATION:  
APPLICANT: Chrysler, Susanna M.S.  
APPLICANT: Sinha, Sukanto  
APPLICANT: Keim, Pamela S.  
APPLICANT: Anderson, John P.  
TITLE OF INVENTION: Beta-Secretase  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Ctr., 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/660,531  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,498  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 15270-002210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2422  
TELEFAX: 415-326-2400  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-660-531-15

Query Match 81.8%; Score 27; DB 4; Length 42;  
Best Local Similarity 85.7%; Pred. No. 4.7;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEE 7  
||| |||  
Db 24 VNLAEE 30

RESULT 15  
US-08-415-751-12  
Sequence 12, Application US/08415751  
Patent No. 5643772

GENERAL INFORMATION:  
APPLICANT: PETERSEN, CAROLYN  
APPLICANT: LEECH, JAMES  
APPLICANT: NELSON, RICHARD, C.  
APPLICANT: GUT, JIRI  
TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-  
TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA  
TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID  
TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND  
TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND  
TITLE OF INVENTION: DIAGNOSIS AND KIT  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PHILLIPS, MOORE, LEMPJO & FINLEY  
STREET: 385 Sherman Avenue, Suite 6  
CITY: Palo Alto  
STATE: California  
COUNTRY: United States of America  
ZIP: 94306-1840  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage  
COMPUTER: PC  
OPERATING SYSTEM: DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/415,751  
FILING DATE: 03-APR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/071,880  
FILING DATE: June 1, 1993  
APPLICATION NUMBER: 07/891,301  
FILING DATE: May 29, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Hana Dolezalova  
REGISTRATION NUMBER: 30,518  
REFERENCE/DOCKET NUMBER: 480.19-2 (HMD)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-1677  
TELEFAX: (415) 324-1678  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 47 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
ORGANISM: Cryptosporidium parvum  
FEATURE:  
NAME/KEY: Positions coded by nonsense codons are  
NAME/KEY: Identified as Xaa.  
US-08-415-751-12

Query Match 81.8%; Score 27; DB 1; Length 47;  
Best Local Similarity 85.7%; Pred. No. 5.4;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VNLAEEF 7  
| | | | |  
DB 7 VELAAEF 13

Search completed: September 6, 2001, 16:39:39  
Job time: 128 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:31 ; Search time 225.25 Seconds  
(without alignments)  
1.884 Million cell updates/sec

Title: US-09-603-713-27

Perfect score: 33  
Sequence: 1 VNLAEEF 7

Scoring table: BLOSUM62

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

```

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

```

```

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

1: A\_genseq\_0601.\*  
2: /SID58/gcgdata/genseq/genseq/A11960.DAT.\*  
3: /SID58/gcgdata/genseq/genseq/A11961.DAT.\*  
4: /SID58/gcgdata/genseq/genseq/A11962.DAT.\*  
5: /SID58/gcgdata/genseq/genseq/A11963.DAT.\*  
6: /SID58/gcgdata/genseq/genseq/A11964.DAT.\*  
7: /SID58/gcgdata/genseq/genseq/A11965.DAT.\*  
8: /SID58/gcgdata/genseq/genseq/A11966.DAT.\*  
9: /SID58/gcgdata/genseq/genseq/A11967.DAT.\*  
10: /SID58/gcgdata/genseq/genseq/A11968.DAT.\*  
11: /SID58/gcgdata/genseq/genseq/A11969.DAT.\*  
12: /SID58/gcgdata/genseq/genseq/A11970.DAT.\*  
13: /SID58/gcgdata/genseq/genseq/A11971.DAT.\*  
14: /SID58/gcgdata/genseq/genseq/A11972.DAT.\*  
15: /SID58/gcgdata/genseq/genseq/A11973.DAT.\*  
16: /SID58/gcgdata/genseq/genseq/A11974.DAT.\*  
17: /SID58/gcgdata/genseq/genseq/A11975.DAT.\*  
18: /SID58/gcgdata/genseq/genseq/A11976.DAT.\*  
19: /SID58/gcgdata/genseq/genseq/A11977.DAT.\*  
20: /SID58/gcgdata/genseq/genseq/A11978.DAT.\*  
21: /SID58/gcgdata/genseq/genseq/A11979.DAT.\*  
22: /SID58/gcgdata/genseq/genseq/A11980.DAT.\*  
23: /SID58/gcgdata/genseq/genseq/A11981.DAT.\*  
24: /SID58/gcgdata/genseq/genseq/A11982.DAT.\*  
25: /SID58/gcgdata/genseq/genseq/A11983.DAT.\*  
26: /SID58/gcgdata/genseq/genseq/A11984.DAT.\*  
27: /SID58/gcgdata/genseq/genseq/A11985.DAT.\*  
28: /SID58/gcgdata/genseq/genseq/A11986.DAT.\*  
29: /SID58/gcgdata/genseq/genseq/A11987.DAT.\*  
30: /SID58/gcgdata/genseq/genseq/A11988.DAT.\*  
31: /SID58/gcgdata/genseq/genseq/A11989.DAT.\*  
32: /SID58/gcgdata/genseq/genseq/A11990.DAT.\*  
33: /SID58/gcgdata/genseq/genseq/A11991.DAT.\*  
34: /SID58/gcgdata/genseq/genseq/A11992.DAT.\*  
35: /SID58/gcgdata/genseq/genseq/A11993.DAT.\*  
36: /SID58/gcgdata/genseq/genseq/A11994.DAT.\*  
37: /SID58/gcgdata/genseq/genseq/A11995.DAT.\*  
38: /SID58/gcgdata/genseq/genseq/A11996.DAT.\*  
39: /SID58/gcgdata/genseq/genseq/A11997.DAT.\*  
40: /SID58/gcgdata/genseq/genseq/A11998.DAT.\*  
41: /SID58/gcgdata/genseq/genseq/A11999.DAT.\*  
42: /SID58/gcgdata/genseq/genseq/A12000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

No.	Score	Query Match	Length	DB	ID	Description
1	33	100.0	7	22	AAB656585	Human memapsin 2
2	33	100.0	7	22	AAB612347	Peptide OM99-1
3	33	100.0	8	22	AAB665686	Human memapsin 2
4	33	100.0	8	22	AAB61348	Peptide OM99-2
5	33	100.0	9	21	AAB07879	A peptide fragment
6	29	87.9	9	21	AAB07885	A peptide fragment
7	29	87.9	14	21	AAB07888	A peptide fragment
8	29	87.9	657	21	AAV52193	Beta glucosidase
9	28	84.8	9	21	AAB07882	A peptide fragment
10	28	84.8	878	22	AAB73265	Yeast trehalase
11	27	81.8	8	21	AAV94771	Beta-secretase sub

12	27	81.8	9	21	AAH82801	Fluoregenic protea
13	27	81.8	9	19	AAH87874	A peptide fragment
14	27	81.8	9	21	AAH87880	A peptide fragment
15	27	81.8	9	21	AAH87894	Substrate for beta
16	27	81.8	10	18	AAH08362	Beta-secretase sub
17	27	81.8	10	20	AAH33756	Synthetic oligopep
18	27	81.8	10	21	AAH69707	Beta-APP alpha-sec
19	27	81.8	10	22	AAH66575	Synthetic peptide
20	27	81.8	10	22	AAH61337	Swedish mutation p
21	27	81.8	16	21	AAH06316	Human beta-amyloid
22	27	81.8	20	19	AAH62211	Fluoregenic protea
23	27	81.8	20	21	AAH69271	Beta-APP alpha-sec
24	27	81.8	21	18	AAH08361	Beta-secretase sub
25	27	81.8	21	19	AAH82166	Fluoregenic protea
26	27	81.8	21	20	AAH33755	Synthetic oligopep
27	27	81.8	30	18	AAH08360	Beta-secretase sub
28	27	81.8	30	20	AAH33754	Synthetic oligopep
29	27	81.8	30	21	AAH07895	Substrate for beta
30	27	81.8	32	17	AAH04402	Mouse amyloid prec
31	27	81.8	32	17	AAH04401	Mouse amyloid prec
32	27	81.8	32	17	AAH04401	Mouse amyloid prec
33	27	81.8	33	18	AAH08359	Beta-secretase sub
34	27	81.8	33	20	AAH33753	Synthetic oligopep
35	27	81.8	33	21	AAH07892	Substrate for beta
36	27	81.8	39	21	AAH69718	Beta-APP alpha-sec
37	27	81.8	42	18	AAH08350	Wild type APP beta
38	27	81.8	42	20	AAH33752	Synthetic oligopep
39	27	81.8	58	20	AAH98001	Swedish-FAD APP714
40	27	81.8	103	16	AAH74698	Beta-amyloid precu
41	27	81.8	115	20	AAH67997	Swedish-FAD APP po
42	27	81.8	115	20	AAH98000	SwedishLondon-FAD
43	27	81.8	317	22	AAH79582	Corynebacterium gl
44	27	81.8	335	21	AAH28726	Arabidopsis thalia
45	27	81.8	394	21	AAH28727	Arabidopsis thalia

## ALIGNMENTS

## RESULT 1

ID	standard; Peptide; 7 AA
AAB66585	

AC AAB66585;

DT 12-APR-2001 (first entry)

Human mapsln 2 inhibitor OM99-1

KW Human memapsin 2; nootropic; neuroprotective; amyloid precursor protein,  $\beta$ -secretase 2 inhibitor; Alzheimer's disease

OS Synthetic

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

[illegible]

PN W0200100665-A2

PD 04-JAN-2001.

PF 27-JUN-2000; 2000WO-US17742  
XY

PR 28-JUN-1999; 99US-0141363;  
PR 30-NOV-1999; 99US-0168060

PK	23-JAN-2000; 2000US-01/1636
PR	27-JAN-2000; 2000US-0178368

XX 06-JUN-2000, 200005 021022Z

PA (UNIT ) UNIV ILLINOIS FOUND

XX Tang JUN, Hong L, Ghosh AK;  
 XX WPI; 2001-137933/14.  
 XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage -  
 PS Example 7; Page 36; 86pp; English.  
 CC The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 CC Sequence 7 AA;  
 SQ  
 Query Match 100.0%; Score 33; DB 22; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3,4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VNLAEPF 7  
 |||||  
 Db 1 vnlaeef 7  
 RESULT 2  
 AAB61347  
 ID AAB61347 standard; peptide; 7 AA.  
 AC AAB61347;  
 XX  
 DT 02-APR-2001 (first entry)  
 XX  
 DE Peptide OM99-1.  
 XX  
 KW Memapsin 2; catalyst; Alzheimer's.  
 OS Unidentified.  
 OS WO200100663-A2.  
 PN  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 27-JUN-2000; 2000WO-US17661.  
 XX  
 PR 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX  
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 XX  
 PI Tang JUN, Lin X, Koelsch G;  
 XX  
 DR WPI; 2001-102885/11.  
 XX  
 PT Purified recombinant catalytically active memapsin 2, used to screen  
 PT inhibitors of it, which are used to treat and prevent Alzheimer's  
 PT disease -  
 PS Example 7; Page 36; 86pp; English.  
 XX  
 CC The present invention relates to a purified recombinant

CC catalytically active memapsin 2. The invention may be used for  
 CC isolating inhibitors which are used to treat or prevent  
 CC Alzheimer's disease. The invention may also be used to screen  
 CC for individuals more genetically prone to develop Alzheimer's  
 CC disease.  
 CC Sequence 7 AA;  
 SQ  
 Query Match 100.0%; Score 33; DB 22; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3,4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VNLAEPF 7  
 |||||  
 Db 1 vnlaeef 7  
 RESULT 3  
 AAB66586  
 ID AAB66586 standard; Peptide; 8 AA.  
 AC AAB66586;  
 XX  
 DT 12-APR-2001 (first entry)  
 XX  
 DE Human memapsin 2 inhibitor OM99-2.  
 XX  
 KW Human memapsin 2; neurotrophic; neuroprotective; amyloid precursor protein;  
 KW APP; memapsin 2 inhibitor; Alzheimer's disease.  
 XX  
 OS Synthetic.  
 OS  
 FT Key Location/Qualifiers  
 FT Modified-site 4..5  
 FT /note="residues 4 and 5 form a Leu-Ala dipeptide  
 FT Isotere"  
 XX  
 PN WO200100665-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 27-JUN-2000; 2000WO-US17742.  
 XX  
 PR 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX  
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 XX  
 PI (UNIT ) UNIV ILLINOIS FOUND.  
 XX  
 PI Tang JUN, Hong L, Ghosh AK;  
 XX  
 DR WPI; 2001-137933/14.  
 XX  
 PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage -  
 PS Example 7; Page 36; 86pp; English.  
 XX  
 CC The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.



XX Sequence 8 AA;

Query Match 100.0%; Score 33; DB 22; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
| | | | | | |  
Db 2 vnlaeef 8

## RESULT 4

AAB61348  
ID AAB61348 standard; peptide; 8 AA.

AC AAB61348;

DT 02-APR-2001 (first entry)

XX Peptide OM99-2.

XX Memapsin 2; catalyst; Alzheimer's.

XX Unidentified.

XX WO200100663-A2.

PD 04-JAN-2001.

PF 27-JUN-2000; 2000WO-US17661.

PR 28-JUN-1999; 99US-0141363.

PR 30-NOV-1999; 99US-0168060.

PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.

PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX PA Tang JUN, Lin X, Koelsch G;

XX PI WPI; 2001-102885/11.

XX DR Purified recombinant catalytically active memapsin 2, used to screen

XX PT inhibitors of it, which are used to treat and prevent Alzheimer's

XX PT disease -

XX Example 7; Page 36; 86pp; English.

XX PS The present invention relates to a purified recombinant

XX CC catalytically active memapsin 2. The invention may be used for

XX CC isolating inhibitors which are used to treat or prevent

XX CC Alzheimer's disease. The invention may also be used to screen

XX CC for individuals more genetically prone to develop Alzheimer's

XX CC disease.

XX SO Sequence 8 AA;

Query Match 100.0%; Score 33; DB 22; Length 8;

Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
| | | | | | |  
Db 2 vnlaeef 8

## RESULT 5

AAB07879  
ID AAB07879 standard; peptide; 9 AA.

XX AAB07879;

XX 14-NOV-2000 (first entry)

XX A peptide fragment derived from beta-amyloid precursor protein.

XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;

XX amyloid plaque component; Alzheimer's disease; amyloidogenic disease;

XX Inhibitor.

XX Homo sapiens.

XX WO200047618-A2.

XX 17-AUG-2000.

XX 10-FEB-2000; 2000WO-US03819.

XX 10-FEB-1999; 99US-0119571.

XX 15-JUN-1999; 99US-0139172.

XX (ELAN-) ELAN PHARM INC.

XX Anderson JP, Basl G, Doane MT, Frigon N, John V, Power M;

XX Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

XX WPI; 2000-533011/48.

XX Purified beta-secretase protein used in assays to discover inhibitors

XX which can be used for the treatment of amyloidogenic diseases e.g.

XX Alzheimer's disease -

XX Disclosure; Page 12; 121pp; English.

XX The specification describes a beta-secretase enzyme. The enzyme cleaves

XX beta-amyloid precursor protein to produce beta-amyloid peptide. This

XX enzyme is therefore implicated in the production of amyloid plaque

XX components which accumulate in the brains of individuals afflicted with

XX Alzheimer's disease. Inhibitors of beta-secretase are administered to

XX a mammalian subject e.g. with Alzheimer's disease or Alzheimer's

XX disease-like pathology to test if they maintain or improve cognitive

XX ability or reduce the plaque burden. The compounds are used for the

XX treatment of amyloidogenic diseases e.g. Alzheimer's disease. The

XX present sequence represents a peptide derived from beta-amyloid

XX precursor protein

XX Sequence 9 AA;

Query Match 100.0%; Score 33; DB 21; Length 9;

Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VNLAEEF 7  
| | | | | | |  
Db 3 vnlaeef 9

## RESULT 6

AAB07885  
ID AAB07885 standard; peptide; 9 AA.

AC AAB07885;

DT 14-NOV-2000 (first entry)

XX A peptide fragment derived from beta-amyloid precursor protein.

XX Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;

XX amyloid plaque component; Alzheimer's disease; amyloidogenic disease;

XX Inhibitor.

OS Homo sapiens.  
 XX MO200047618-A2.  
 PN 17-AUG-2000.  
 PD 10-FEB-2000; 2000WO-US03819.  
 PF 10-FEB-1999; 99US-0119571.  
 PR 15-JUN-1999; 99US-0139172.  
 XX (ELAN-) ELAN PHARM INC.  
 PA Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
 PI MPI; 2000-533011/48.  
 DR Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease -  
 PS Disclosure; Page 12; 121pp; English.  
 CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents a peptide derived from beta-amyloid  
 CC precursor protein  
 XX Sequence 9 AA:  
 SQ

Query Match 87.9%; Score 29; DB 21; Length 9;  
 Best Local Similarity 85.7%; Pred. No. 3.4e+05;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAER 7  
 II IIII  
 Db 3 vnlaerf 9

RESULT 7  
 AAB07888  
 ID AAB07888 standard; peptide; 14 AA.  
 XX  
 AC AAB07888;  
 XX  
 DT 14-NOV-2000 (first entry)  
 XX  
 DE A peptide fragment derived from beta-amyloid precursor protein.  
 XX  
 KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 KM inhibitor.  
 XX  
 XX Homo sapiens.  
 OS  
 PN WO200047618-A2.  
 PD 17-AUG-2000.  
 PF 10-FEB-2000; 2000WO-US03819.  
 PR 10-FEB-1999; 99US-0119571.  
 PR 15-JUN-1999; 99US-0139172.  
 XX

PA (ELAN-) ELAN PHARM INC.  
 XX  
 PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
 PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
 XX MPI; 2000-533011/48.  
 DR Purified beta-secretase protein used in assays to discover inhibitors  
 PT which can be used for the treatment of amyloidogenic diseases e.g.  
 PT Alzheimer's disease -  
 PS Disclosure; Page 12; 121pp; English.  
 CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
 CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
 CC enzyme is therefore implicated in the production of amyloid plaque  
 CC components which accumulate in the brains of individuals afflicted with  
 CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
 CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
 CC disease-like pathology to test if they maintain or improve cognitive  
 CC ability or reduce the plaque burden. The compounds are used for the  
 CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
 CC present sequence represents a peptide derived from beta-amyloid  
 CC precursor protein  
 XX Sequence 14 AA:  
 SQ

Query Match 87.9%; Score 29; DB 21; Length 14;  
 Best Local Similarity 85.7%; Pred. No. 1.2;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAER 7  
 III IIII  
 Db 8 vnlaerf 14

RESULT 8  
 ID AAY52193 standard; Protein; 657 AA.  
 XX  
 AC AAY52193;  
 XX  
 DT 15-FEB-2000 (first entry)  
 XX  
 DE Beta glucosidase (Bgl A) amino acid sequence.  
 XX  
 KW Bgl A; beta glucosidase; cellulose; cellulosic biomass.  
 XX  
 OS Orpinomyces sp.  
 XX  
 FT Key Location/Qualifiers  
 FT Peptide 1..16  
 FT Protein /note="Putative signal peptide"  
 FT /note="Beta-glucosidase"  
 XX  
 PN WO9957287-A2.  
 XX  
 PD 11-NOV-1999.  
 PF 06-MAY-1999; 99WO-US10106.  
 PR 06-MAY-1998; 98US-0084494.  
 XX  
 XX (UYGE-) UNIV GEORGIA RES FOUND INC.  
 PA Li X, Ljungdahl LG, Chen H, Ximenes EA;  
 PI MPI; 2000-023580/02.  
 DR N-PSDB; AA237865.  
 XX  
 PT New orpinomyces pc-2 beta-glucosidase coding sequences and protein

XX Claim 1; Page 35-36; 63pp; English.  
PS  
XX  
CC This is the beta-glucosidase (Bg1a) amino acid sequence from *Orpinomyces*  
CC sp. strain PC-2. Beta-glucosidase acts to liberate D-glucose units from  
CC cellobiose, cello-oligosaccharides and other glucosides. A host cell  
CC containing the Bg1a DNA sequence can be used to produce a  
CC beta-glucosidase in a host cell other than *Orpinomyces* sp. strain PC-2.  
CC The fungal beta-glucosidase is useful for the improved saccharification  
CC of cellulosic materials. Effective utilization of cellulosic biomass  
CC through biological processes is one approach to overcoming the shortage  
CC of food, feeds and fuels, expected as a consequence of the explosive  
CC increase in human population.  
XX  
SQ Sequence 657 AA:  
  
Query Match 87.9%; Score 29; DB 21; Length 657;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 NLAAEF 7  
| | | | | | |  
DB 377 nlaaef 382  
  
RESULT 9  
AAB07882  
ID AAB07882 standard; peptide; 9 AA.  
XX  
AC AAB07882;  
XX  
DF 14-NOV-2000 (first entry)  
XX  
DE A peptide fragment derived from beta-amyloid precursor protein.  
XX  
KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KW inhibitor.  
XX  
OS Homo sapiens.  
XX  
PN WO20047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PF 10-FEB-2000; 2000MO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
PR 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Baai G, Doane MT, Frigon N, John V, Power M,  
PI Simha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
DR WPI: 2000-533011/48.  
XX  
PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.  
PT Alzheimer's disease -  
XX  
XX  
PS Disclosure; Page 12; 121pp; English.  
XX  
CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The

CC present sequence represents a peptide derived from beta-amyloid  
CC precursor protein  
XX  
SQ Sequence 9 AA:  
  
Query Match 84.8%; Score 28; DB 21; Length 9;  
Best Local Similarity 85.7%; Pred. No. 3,4e+05;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 VNLAAEF 7  
| | | | | | |  
DB 3 vnlaaef 9  
  
RESULT 10  
AAB73265  
ID AAB73265 standard; protein; 878 AA.  
XX  
AC AAB73265;  
XX  
DF 17-MAY-2001 (first entry)  
XX  
DE Yeast trehalose-6-phosphate phosphatase TPP.  
XX  
KW Yeast; TPP; antiparasitic; fungicide; bactericide; protozoacide;  
KW insecticide; trehalose-6-phosphate phosphatase.  
XX  
OS Candida albicans.  
XX  
PN WO200116357-A2.  
XX  
PD 08-MAR-2001.  
XX  
PF 29-AUG-2000; 2000MO-EP08410.  
XX  
PR 30-AUG-1999; 99EP-0202805.  
PR 27-JUN-2000; 2000EP-0870145.  
XX  
PA (LEUV-) LEUVEN RES & DEV.  
XX  
PI Thevelein J, Van Dijk P;  
XX  
DR WPI: 2001-235118/24.  
XX  
PT New inhibitors and screening assay for inhibitors or suppressors of  
PT sugar alcohol phosphatases or sugar phosphatases -  
XX  
PS Disclosure; Fig 8; 106pp; English.  
XX  
CC The present invention relates to a screening assay for inhibitors or  
CC suppressors of sugar alcohol phosphatases or sugar phosphatases using  
CC enzymes found in fungi, bacteria, insects, nematodes, mites or protozoa.  
CC The inhibitors identified are useful for treating or preventing  
CC parasitic, particularly fungal, bacterial or protozoal infections, or a  
CC nematode, insect, worm or mite infestation, in a human, animal or plant.  
CC The present sequence is one such sugar phosphatase: yeast  
CC trehalose-6-phosphate phosphatase (TPP).  
XX  
SQ Sequence 878 AA:  
  
Query Match 84.8%; Score 28; DB 22; Length 878;  
Best Local Similarity 85.7%; Pred. No. 2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 VNLAAEF 7  
| | | | | | |  
DB 634 vnlaaef 640  
  
RESULT 11  
AAV94771

```

ID  AAY94771 standard; Protein: 8 AA.
XX
AC  AAY94771;
XX
DT  12-FEB-2001 (first entry)
XX
DE  Beta-secretase substrate peptide SEQ ID 17.
XX
XX  Beta-secretase; enzyme; amyloid plaque; Alzheimer's disease;
KM  Down's syndrome; amyloid angiopathy; gene therapy; neuroprotective.
XX
OS  Synthetic.
XX
PN  WO200058479-A1.
XX
PD  05-OCT-2000.
XX
PF  23-MAR-2000; 2000MO-US07755.
XX
PR  26-MAR-1999; 99US-0277229.
XX
PA  (AMGE-) AMGEN INC.
XX
PI  Citron M, Vassar RJ, Bennett BD;
XX
DR  WPI; 2000-594643/56.
XX
PT  Isolated beta-secretase nucleic acids and encoded polypeptides, useful
PS  for diagnosis and gene therapy of Alzheimer's disease -
XX
XX  Example 10; Page 117; 145pp; English.
XX
CC  This invention relates to 3 nucleotide sequences encoding beta-secretase
CC  proteins. Beta-secretase is an enzyme involved in the production of one
CC  of the components of amyloid plaques involved in Alzheimer's disease. The
CC  invention includes an expression vector comprising the nucleotide
CC  sequence, a host cell comprising the expression vector, and a process for
CC  producing the protein through culturing the transformed cells. Also
CC  included in the invention are a polypeptide derivative of the
CC  beta-secretase protein, a fusion protein comprising beta-secretase fused
CC  to a heterologous amino acid sequence, and a method for modulating the
CC  levels of beta-secretase polypeptide in a mammal comprising administering
CC  the polynucleotide sequence. Beta-secretase exhibits neuroprotective and
CC  neurotropic activity. The beta-secretase gene and related genes on chromosomes
CC  map locations of the beta-secretase gene and related genes on chromosomes
CC  and as hybridization probes in diagnostic assays to test for the presence
CC  of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's
CC  syndrome, and amyloid angiopathy. The nucleotide sequence may also be
CC  used as anti-sense inhibitors of beta-secretase expression, in gene
CC  therapy of Alzheimer's disease, and for the identification of compounds
CC  that modulate beta-secretase activity. Antibodies to the beta-secretase
CC  protein may be used for in vitro and in vivo diagnostic purposes to
CC  detect the presence of beta-secretase polypeptide in a body fluid or cell
CC  sample. The present sequence represents a beta-secretase substrate
CC  peptide.
XX
SQ  Sequence 8 AA:

```

Query Match 81.8%; Score 27; DB 21; Length 8;  
 Best Local Similarity 85.7%; Pred. No. 3.4e+05;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY  1 VNLADEF 7
    ||| |||
DB  2 vnlaef 8

```

RESULT 12  
 ID AAW82081 standard; peptide; 9 AA.  
 XX  
 XX AAW82081;

```

XX
DT  18-FEB-1999 (first entry).
XX
XX  Fluorogenic protease indicator protease binding peptide #59.
DE  Protease activity; fluorophore; detection; fluorogenic; cellular uptake;
XX  conformation change.
XX
OS  Synthetic.
XX
PN  WO9837226-A1.
XX
PD  27-AUG-1998.
XX
PF  20-FEB-1998; 98MO-US03000.
XX
PR  20-FEB-1997; 97US-0802981.
XX
PA  (ONCO-) ONCOMMUNIN INC.
XX
PI  Komoriya A, Packard BS;
XX
DR  WPI; 1998-467579/40.
XX
PT  New fluorogenic compositions - containing 2 fluorophores separated
PS  by a peptide comprising a protease binding site, used for detecting
XX  protease activity in samples.
XX
XX  Claim 4; Page 77; 90pp; English.
XX
CC  AAW82023-W82240 are peptides used in the construction of a fluorogenic
CC  composition which is used for the detection of protease activity in
CC  biological samples. The products can be used for the detection of
CC  conformation changes in nucleic acids, oligosaccharides,
CC  polysaccharides, proteins, peptides, lipids, phospholipids, glycolipids,
CC  glycoproteins, steroids or polymers. In addition, attachment of a
CC  hydrophobic group to a molecule can be used to enhance uptake by cells.
CC  The composition is composed of P = peptide comprising a protease binding
CC  site for the protease, F1, F2 peptides - fluorophores where F1 is
CC  attached to the amino terminal amino acid and F2 is attached to the
CC  carboxyl terminal amino acid and S1, S2 peptides - when present, are
CC  peptide spacers where S1, when present, is attached to the amino terminal
CC  acid, and S2, when present, is attached to the carboxyl terminal amino
CC  acid.
XX
SQ  Sequence 9 AA:

```

Query Match 81.8%; Score 27; DB 19; Length 9;  
 Best Local Similarity 85.7%; Pred. No. 3.4e+05;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY  1 VNLADEF 7
    ||| |||
DB  3 vnlaef 9

```

RESULT 13  
 ID AAB07874 standard; peptide; 9 AA.  
 XX  
 XX AAB07874;  
 XX  
 DT 14-NOV-2000 (first entry)  
 XX  
 XX A peptide fragment derived from beta-amyloid precursor protein.  
 DE Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
 KM amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
 KM inhibitor.  
 XX  
 OS Homo sapiens.  
 XX

PN WO20047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PE 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
XX 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
DR WPI: 2000-533011/48.  
XX  
PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.  
PT Alzheimer's disease -  
XX  
PS Disclosure: Page 12; 121pp; English.  
XX  
CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a peptide derived from beta-amyloid  
CC precursor protein  
XX  
SQ Sequence 9 AA:  
  
Query Match 81.8%; Score 27; DB 21; Length 9;  
Best Local Similarity 85.7%; Pred. NO. 3.4e+05;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 VNLAEEF 7  
DB 3 VNLDAEF 9  
  
RESULT 14  
AAB07880  
ID AAB07880 standard; peptide: 9 AA.  
XX  
AC AAB07880;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE A peptide fragment derived from beta-amyloid precursor protein.  
XX  
KM Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KM amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KM inhibitor.  
XX  
OS Homo sapiens.  
XX  
PN WO20047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PE 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
XX 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX

PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
DR WPI: 2000-533011/48.  
XX  
PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.  
PT Alzheimer's disease -  
XX  
PS Disclosure: Page 12; 121pp; English.  
XX  
CC The specification describes a beta-secretase enzyme. The enzyme cleaves  
CC beta-amyloid precursor protein to produce beta-amyloid peptide. This  
CC enzyme is therefore implicated in the production of amyloid plaque  
CC components which accumulate in the brains of individuals afflicted with  
CC Alzheimer's disease. Inhibitors of beta-secretase are administered to  
CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
CC disease-like pathology to test if they maintain or improve cognitive  
CC ability or reduce the plaque burden. The compounds are used for the  
CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
CC present sequence represents a peptide derived from beta-amyloid  
CC precursor protein  
XX  
SQ Sequence 9 AA:  
  
Query Match 81.8%; Score 27; DB 21; Length 9;  
Best Local Similarity 85.7%; Pred. NO. 3.4e+05;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 VNLAEEF 7  
DB 3 VNLDAEF 9  
  
RESULT 15  
AAB07894  
ID AAB07894 standard; Peptide: 9 AA.  
XX  
AC AAB07894;  
XX  
DT 14-NOV-2000 (first entry)  
XX  
DE Substrate for beta-secretase enzyme.  
XX  
KM Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;  
KM amyloid plaque component; Alzheimer's disease; amyloidogenic disease;  
KM inhibitor.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Cleavage-site 5..6  
XX  
PN WO20047618-A2.  
XX  
PD 17-AUG-2000.  
XX  
PE 10-FEB-2000; 2000WO-US03819.  
XX  
PR 10-FEB-1999; 99US-0119571.  
XX 15-JUN-1999; 99US-0139172.  
XX  
PA (ELAN-) ELAN PHARM INC.  
XX  
PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;  
PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;  
XX  
DR WPI: 2000-533011/48.  
XX  
PT Purified beta-secretase protein used in assays to discover inhibitors  
PT which can be used for the treatment of amyloidogenic diseases e.g.  
PT Alzheimer's disease -

xx Example 4: Page 71; 121pp; English.

ps  
xx  
cc The specification describes a beta-secretase enzyme. The enzyme cleaves  
cc beta-amyloid precursor protein to produce beta-amyloid peptide. This  
cc enzyme is therefore implicated in the production of amyloid plaque  
cc components which accumulate in the brains of individuals afflicted with  
cc Alzheimer's disease. Inhibitors of beta-secretase are administered to  
cc a mammalian subject e.g. with Alzheimer's disease or Alzheimer's  
cc disease-like pathology to test if they maintain or improve cognitive  
cc ability or reduce the plaque burden. The compounds are used for the  
cc treatment of amyloidogenic diseases e.g. Alzheimer's disease. The  
cc present sequence represents a peptide substrate used to test the  
cc activity of beta-secretase enzyme.

xx  
sq Sequence 9 AA;

Query Match 81.8%; Score 27; DB 21; Length 9;  
Best Local Similarity 85.7%; Pred. NO. 3.4e+05;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VNLAEF 7  
||| |||  
Db 3 vnldeaf 9

Search completed: September 6, 2001, 16:43:32  
Job time: 361 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:54 ; Search time 231.42 Seconds  
(without alignments)  
4.002 Million cell updates/sec

Title: US-09-603-713-26

Perfect score: 32

Sequence: 1 GVLSRK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 125026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_16:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_unclassified:\*  
14: sp\_vertebrate:\*  
15: sp\_virus:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	93.8	553	14	Q08541 sugar beet
2	30	93.8	553	14	Q06110 beet wester
3	29	90.6	98	8	O63589 florometra
4	29	90.6	204	14	O83099 latino vtru
5	29	90.6	476	2	O50017 mycobacteri
6	29	90.6	558	14	Q84169 oliveros vi
7	29	90.6	1194	13	Q9W737 gallus gall
8	28	87.5	301	3	O94631 schizosacch
9	28	87.5	451	4	O9H9K4 homo sapien
10	28	87.5	464	4	O9P249 homo sapien
11	28	87.5	464	4	O9H9G5 homo sapien
12	28	87.5	487	2	O9L3L1 salmonella
13	28	87.5	534	3	O9HDW1 schizosacch
14	28	87.5	670	10	O9FF80 arabidopsids
15	28	87.5	719	10	O9M399 arabidopsids
16	28	87.5	1060	5	O9VSR1 drosophila
17	27	84.4	87	8	O9G8S4 naegleria g
18	27	84.4	182	13	O9W663 trachemys s
19	27	84.4	182	13	O9W661 trachemys s

20	27	84.4	182	13	O9W660 poephila gu
21	27	84.4	182	13	O9W659 poephila gu
22	27	84.4	182	13	O9W657 coryphaenoi
23	27	84.4	182	13	O9W655 carassius a
24	27	84.4	182	13	O9W6K5 brachydanto
25	27	84.4	182	13	O9W6P6 alepocephal
26	27	84.4	182	13	O9W6P4 lophius pis
27	27	84.4	182	13	O9W6P2 rana pipien
28	27	84.4	182	13	O9W6P1 chelydra se
29	27	84.4	182	13	O9W6P0 chelydra se
30	27	84.4	182	13	O9W6E9 xenopus lae
31	27	84.4	185	2	O9W6N16 helicobacte
32	27	84.4	196	2	O9W6YF6 delnococcus
33	27	84.4	204	1	O9W6SG4 halobacteri
34	27	84.4	232	13	O9J276 brachydanto
35	27	84.4	248	7	O9W6Q44 aulonocara
36	27	84.4	282	13	O9W6D11 gallus gall
37	27	84.4	341	4	O9W6D87 homo sapien
38	27	84.4	366	1	O9W6V124 pyrococcus
39	27	84.4	387	10	O9W6LNP0 arabidopsids
40	27	84.4	392	2	O9W6RAJ9 moraxella s
41	27	84.4	397	1	O58488 pyrococcus
42	27	84.4	419	4	O9W6UC15 homo sapien
43	27	84.4	436	10	O65215 zea mays (m
44	27	84.4	471	2	O06959 vibrio chol
45	27	84.4	480	3	O9W6783 schizosacch

## ALIGNMENTS

RESULT 1  
008541 PRELIMINARY: PRT; 553 AA.  
AC Q08541;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE HEAT SHOCK PROTEIN 90 HOMOLOG.  
OS Sugar beet yellow virus (SBVY).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Closteroviridae;  
OC Closterovirus.  
OX NCBI\_TaxID=31714;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPRAIN-UKRAINIAN.  
RX MEDLINE=94082464; PubMed=8259666;  
RA Agratovsky A.A., Koonin E.V., Boyko V.P., Maiss E., Froetschl R.,  
RA Lunina N.A., Alabekov J.G.;  
RT "Beet yellows closterovirus: complete genome structure and  
RT identification of a leader papain-like thiol protease.\*";  
RL Virology 198:311-324(1994).  
CC -1- SIMILARITY: BELONGS TO HEAT SHOCK PROTEIN 90 FAMILY.  
DR EMBL, X73476; CAA51866.1; -.  
KW Heat shock.  
SQ SEQUENCE 553 AA; 63904 MW; E7811795E2681804 CRC64;

Query Match 93.8%; Score 30; DB 14; Length 553;  
Best Local Similarity 85.7%; Pred. No. 80;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
||:||||  
Db 50 GVLSRK 56

RESULT 2  
066110 PRELIMINARY: PRT; 553 AA.  
ID 066110;  
AC 066110;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMblrel. 08, last annotation update)  
 DE BEET YELLOW VIRUS GENOME 3'-PROXIMAL HALF, FOR CAPSID PROTEIN HSP70  
 OS RELATED PROTEIN, RNA-DEPENDENT RNA POLYMERASE AND ORFS.  
 OC Beet western yellows virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
 OC Polerovirus.  
 OX NCBI\_TaxID=12042;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Agronomovsky A.A.;  
 RL Submitted (MAY-1990) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Agronomovsky A.A.;  
 RL Submitted (JUL-1990) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=91116305; PubMed=1990061;  
 RA Agronomovsky A.A., Boyko V.P., Karasev A.V., Lunina N.A., Koonin E.V.,  
 RA Dolja V.V.;  
 RT "Nucleotide sequence of the 3'-terminal half of beet yellows  
 RT closterovirus RNA genome: unique arrangement of eight virus genes.";  
 RL J. Gen. Virol. 72:15-23(1991).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=91171285; PubMed=2005613;  
 RA Agronomovsky A.A., Boyko V.P., Karasev A.V., Koonin E.V., Dolja V.V.;  
 RT "Putative 65 kDa protein of beet yellows closterovirus is a homologue  
 RT of HSP70 heat shock proteins";  
 RL J. Mol. Biol. 217:603-610(1991).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94343547; PubMed=7545947;  
 RA Rensing S.A., Maier U.G.;  
 RT "Phylogenetic analysis of the stress-70 protein family.";  
 RL J. Mol. Evol. 39:80-86(1984).  
 DR EMBL; X53462; CAA37552.1; -;  
 SQ SEQUENCE 553 AA; 64004 MW; E7940295E26802B4 CRC64;

Query Match 93.8%; Score 30; DB 14; Length 553;  
 Best Local Similarity 85.7%; Pred. No. 80;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 ||:||||  
 DB 50 GVLRSK 56

RESULT 3  
 ID 063589 PRELIMINARY; PRT; 98 AA.  
 AC 063589;  
 DT 01-AUG-1998 (TREMblrel. 07, Created)  
 DT 01-AUG-1998 (TREMblrel. 07, last sequence update)  
 DT 01-OCT-2000 (TREMblrel. 15, last annotation update)  
 DE NADH DEHYDROGENASE SUBUNIT 4L.  
 OS Floumetra serratissima.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Echinodermata; Pelmatozoa; Crinoidea; Articulata;  
 OC Comatulida; Antedonidae; Floumetra.  
 OX NCBI\_TaxID=73431;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Scouras A., Smith M.J.;  
 RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF049132; AAD05075.1; -;  
 DR InterPro; IPR001133; -;  
 DR InterPro; IPR003214; -;  
 DR Pfam; PF00420; oxidored\_q2; 1.  
 DR Prodom; PD000359; -; 1.  
 DR Mitochondrion.  
 DR Mitochondrion.  
 SQ SEQUENCE 98 AA; 10958 MW; 64B109235792A2F0 CRC64;

Query Match 90.6%; Score 29; DB 8; Length 98;  
 Best Local Similarity 85.7%; Pred. No. 25;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 ||||:|  
 DB 18 GVLRSK 24

RESULT 4  
 ID 083099 PRELIMINARY; PRT; 204 AA.  
 AC 083099;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 13, last annotation update)  
 DE NUCLEOCAPSID PROTEIN (FRAGMENT).  
 OS Latino virus.  
 OC Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.  
 OX NCBI\_TaxID=45221;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=10924;  
 RA MEDLINE=96204598; PubMed=8623541;  
 RA Bowen M.D., Peters C.J., Nichol S.T.;  
 RT "The phylogeny of New World (Tscaribe complex) arenaviruses.";  
 RL Virology 219:285-290(1996).  
 DR EMBL; U43688; AAC54831.1; -;  
 DR InterPro; IPR000229; -;  
 DR Pfam; PF00843; Arena\_nucleocap; 1.  
 DR Nucleocapsid.  
 DR NON\_TER 1 1  
 FT NON\_TER 204 204  
 SQ SEQUENCE 204 AA; 22690 MW; 9F0D63CBDD80D756 CRC64;

Query Match 90.6%; Score 29; DB 14; Length 204;  
 Best Local Similarity 85.7%; Pred. No. 51;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 ||:||||  
 DB 189 GVLRSK 195

RESULT 5  
 ID 050017 PRELIMINARY; PRT; 476 AA.  
 AC 050017;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
 DT 01-MAY-2000 (TREMblrel. 13, last annotation update)  
 DE XCIC.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith D.R.;  
 RL Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; U15181; AAA62961.1; -;  
 DR HSP; P08659; IIC1.  
 DR InterPro; IPR000873; -;  
 DR InterPro; IPR003015; -;  
 DR Pfam; PF00501; AMP-binding; 1.  
 DR PRINTS; PR00154; AMPBINDING.



DR PROSITE: PS00455; AMP\_BINDING: 1.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX: UNKNOWN\_1.  
 SQ SEQUENCE 476 AA; 50130 MW; 4F8DDED72059E6E6 CRC64;

Query Match 90.6%; Score 29; DB 2; Length 476;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 |||||:  
 DB 149 GVLRSR 155

## RESULT 6

084169 PRELIMINARY: PRT: 558 AA.  
 AC 084169;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE NUCLEOCAPSID PROTEIN.  
 OS Oliveros virus.  
 OC Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.  
 OX NCBI\_TaxID=42764;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3229-1;  
 RX MEDLINE=96177154; PubMed=8599223;  
 RA Bowen M.D., Peters C.J., Mills J.N., Nichol S.T.;  
 RT "Oliveros virus: a novel arenavirus from Argentina."  
 RL Virology 217:362-366(1996).  
 DR EMBL: U34248; AACS4655.1; -;  
 DR InterPro: IPR000229; -;  
 DR InterPro: IPR001998; -;  
 DR Pfam: PF00843; Arena\_nucleocap. 1.  
 DR ProDom: PD004728; -; 1.  
 DR ProSite: PS00172; XYLOSE\_ISOMERASE\_1; UNKNOWN\_1.  
 DR Nucleocapsid.  
 KW Nucleocapsid.  
 SQ SEQUENCE 558 AA; 62154 MW; 3F617AA8396A3D7 CRC64;

Query Match 90.6%; Score 29; DB 14; Length 558;  
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 |||||:  
 DB 502 GVLRSR 508

## RESULT 7

09W737 PRELIMINARY: PRT: 1194 AA.  
 AC 09W737;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE NOTCH-1 (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=99328644; PubMed=10402194;  
 RA Wakamatsu Y., Maynard T.M., Jones S.U., Weston J.A.;  
 RT "NUMB localizes in the basal cortex of mitotic avian neuroepithelial  
 cells and modulates neuronal differentiation by binding to NOTCH-1."  
 RL Neuron 23:71-81(1999).  
 DR EMBL: AF159231; AADA2893.1; -;

DR HSP: P00740; IEDM.  
 DR InterPro: IPR000561; -;  
 DR InterPro: IPR000800; -;  
 DR InterPro: IPR001064; -;  
 DR InterPro: IPR002110; -;  
 DR InterPro: IPR003571; -;  
 DR Pfam: PF00023; ank; 6.  
 DR Pfam: PF00066; notch; 3.  
 DR ProDom: PD000206; -; 1.  
 DR ProDom: PS50088; ANK\_REPEAT; 4.  
 DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE: PS00225; CRYSTALLIN\_BETAGAMMA; UNKNOWN\_1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
 DR SMART: SM00248; ANK; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 1194 AA; 130640 MW; 650380B8E584974 CRC64;

Query Match 90.6%; Score 29; DB 13; Length 1194;  
 Best Local Similarity 85.7%; Pred. No. 3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 |||||:  
 DB 395 GVLRSR 401

## RESULT 8

094631 PRELIMINARY: PRT: 301 AA.  
 AC 094631;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE PUTATIVE RIBOSE METHYLTRANSFERASE C1347.13C (Ec 2.1.1.-).  
 GN SPBC1347.13C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Wood V., Rajandream M.A., Barrell B.G., Rieger M.;  
 RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL (BY SIMILARITY).  
 CC -1- SIMILARITY: TO YEAST PET56.  
 DR EMBL: AL035548; CAB37444.1; -;  
 DR InterPro: IPR001537; -;  
 DR Pfam: PF00588; SPOU\_methylase; 1.  
 DR ProDom: PD001243; -; 1.  
 DR Hypothetical protein; Mitochondrion; Transferase; Methyltransferase.  
 FT DOMAIN 139  
 SQ SEQUENCE 301 AA; 33402 MW; 437087BCA2D7E5B CRC64;

Query Match 87.5%; Score 28; DB 3; Length 301;  
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 |||||:  
 DB 175 GVLRSK 181

## RESULT 9

09H9K4 PRELIMINARY: PRT: 451 AA.  
 ID 09H9K4;  
 AC 09H9K4;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE CDNA FLJ12685 FIS, CLONE NT2RM4002493.  
 OC Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Morikawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Makatsus A., Nakamura Y., Nagahari K., Masuh Y.,  
 RA Niimura Y., Iwayanagi T.,  
 RA "NEO human cDNA sequencing project."  
 RT Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
 RL EMBL: AK022747; BAB14221.1;  
 DR EMBL: AK022747; BAB14221.1;  
 SO SEQUENCE 451 AA; 49681 MW; EA8BFFFE7067AB04 CRC64;

Query Match 87.5%; Score 28; DB 4; Length 451;  
 Best Local Similarity 85.7%; Pred. No. 2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 Db 364 GVLRSQ 370

RESULT 10  
 ID 09P249 PRELIMINARY; PRT; 464 AA.

AC 09P249: 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE KIAA1479. PROTEIN (FRAGMENT).  
 GN KIAA1479.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20277482; PubMed=10819331;  
 RA Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.,  
 RT "Prediction of the coding sequences of unidentified human  
 RT genes. XVII. The complete sequences of 100 new cDNA clones from brain  
 RT which code for large proteins in vitro."  
 RL DNA Res. 7:143-150(2000).  
 DR EMBL: AB040912; BAA96003.1;  
 FT NON\_TER 1  
 SO SEQUENCE 464 AA; 51163 MW; D75AA643B6688DC8 CRC64;

Query Match 87.5%; Score 28; DB 4; Length 464;  
 Best Local Similarity 85.7%; Pred. No. 2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 Db 377 GVLRSQ 383

RESULT 11  
 ID 09H9G5 PRELIMINARY; PRT; 464 AA.

DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE CDNA FLJ12769 FIS, CLONE NT2RP2001581.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Makatsus A., Nakamura Y., Nagahari K., Masuh Y., Oshima A.,  
 RA "NEO human cDNA sequencing project."  
 RT Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AK022831; BAB14264.1;  
 SO SEQUENCE 464 AA; 51214 MW; C850600BAE9A0C94 CRC64;

Query Match 87.5%; Score 28; DB 4; Length 464;  
 Best Local Similarity 85.7%; Pred. No. 2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 Db 377 GVLRSQ 383

RESULT 12  
 ID 09L3L1 PRELIMINARY; PRT; 487 AA.

AC 09L3L1: 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE PHOQ PROTEIN.  
 GN PHOQ.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SL344;  
 RA Cano D.A., Martinez-Moya M., Casadesus J., Groisman E.A.,  
 RT "Attenuation of Salmonella proliferation within host cells mediated by  
 RT pathogen virulence regulators."  
 RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE  
 CC KINASES.  
 DR EMBL: AJ272210; CAB75592.1;  
 DR InterPro: IPR000410;  
 DR InterPro: IPR000658;  
 DR InterPro: IPR003594;  
 DR Pfam: PF00512; signal; 1.  
 DR Pfam: PF00672; DUF5; 1.  
 DR PRINTS: PR00344; BCTRLSENSOR.  
 DR SMART: SM00387; HATPase\_C; 1.  
 KW Kinase; Phosphorylation; Sensory transduction; Transferase.  
 SO SEQUENCE 487 AA; 55466 MW; BDCFEFC56P4CA058 CRC64;

Query Match 87.5%; Score 28; DB 2; Length 487;  
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 Db 331 GVLRSQ 337

RESULT 13  
 ID 09HDW1 PRELIMINARY; PRT; 534 AA.

AC 09HDW1:

DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)  
 DE PUTATIVE PHOSPHATIDYLGLYCEROPHOSPHATE SYNTHASE, THE FIRST ENZYME OF  
 DE THE CARDIOLIPIN BIOSYNTHETIC PATHWAY.  
 GN SPBP18G5.02.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972H-;  
 RA Collins M., Harris D., Wood V., Rajandream M.A., Barrell B.G.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL512563; CAC21490.1; -;  
 SQ SEQUENCE 534 AA; 61428 MW; 1027C2573FACDBCD CRC64;

Query Match 87.5%; Score 28; DB 3; Length 534;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
 I I I I I  
 DB 324 GTLSRK 330

RESULT 14  
 O9FF80 PRELIMINARY; PRT; 670 AA.  
 AC O9FF80;  
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)  
 DE SET-DOMAIN PROTEIN-LIKE.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=COLUMBIA;  
 RX MEDLINE=97471969; Pubmed=9330910;  
 RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
 RA Miyajima N., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
 RT features of the 1.6 Mb regions covered by twenty physically assigned  
 RT clones."  
 RL DNA Res. 4:215-230(1997).  
 DR EMBL: AB005245; BAB11516.1; -;  
 SQ SEQUENCE 670 AA; 74471 MW; C92CE89FF5C630F1 CRC64;

Query Match 87.5%; Score 28; DB 10; Length 670;  
 Best Local Similarity 71.4%; Pred. No. 2.9e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
 I I I I I  
 DB 463 GILVSRK 469

RESULT 15  
 O9M399 PRELIMINARY; PRT; 719 AA.  
 AC O9M399;  
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)  
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)  
 DE HYPOTHETICAL 80.6 KDA PROTEIN.

GN F2K15.260.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,  
 RA Mewes H.W., Lemcke K., Mayer K.F.X., Queller F., Salanoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL132956; CAB66419.1; -;  
 DR InterPro: IPR001680; -;  
 DR Pfam: PF00400; WD40; 5.  
 KW Hypothetical protein.  
 SQ SEQUENCE 719 AA; 80589 MW; 32ED6E5A4652205 CRC64;

Query Match 87.5%; Score 28; DB 10; Length 719;  
 Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
 I I I I I  
 DB 702 GTLSRK 708

Search completed: September 6, 2001, 16:49:55  
 Job time: 739 sec

Fri Sep 7 10:58:05 2001

us-09-603-713-26.rspt

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:10 : Search time 72.75 Seconds

(without alignments)  
3.296 Million cell updates/sec

Title: US-09-603-713-26

Perfect score: 32

Sequence: 1 GVLLSRK 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	2444	1 NTCL_HUMAN	P46531 homo sapien
2	32	100.0	2531	1 NTCL_MOUSE	Q01705 mus musculu
3	32	100.0	2531	1 NTCL_RAT	Q07008 rattus norv
4	29	90.6	101	1 YQUD_ECOLI	P42617 escherichia
5	29	90.6	123	1 RK19_PORPU	P51331 porphyra pu
6	29	90.6	258	1 LIDR_ECOLI	P33233 escherichia
7	28	87.5	422	1 EXUT_BACSU	Q34456 bacillus su
8	28	87.5	431	1 PHOQ_SALTY	P14147 salmoneila
9	28	87.5	786	1 YDS9_SCHPO	Q10313 schizosacch
10	28	87.5	1060	1 VP2_AHSVA	P32253 african hor
11	27	84.4	115	1 RL19_STRFR	Q34031 streptococc
12	27	84.4	117	1 ARRL_ECOLI	P15905 escherichia
13	27	84.4	263	1 CB1J_SALTY	Q05591 salmoneila
14	27	84.4	281	1 RM30_YEAST	P36528 saccharomyc
15	27	84.4	319	1 SAL2_CAEEL	Q20410 caenorhabdi
16	27	84.4	448	1 YN06_YEAST	P53890 saccharomyc
17	27	84.4	457	1 FTSZ_PORGI	Q08466 porphyromon
18	27	84.4	585	1 DCE2_HUMAN	Q05329 homo sapien
19	27	84.4	585	1 DCE2_MOUSE	P48320 mus musculu
20	27	84.4	585	1 DCE2_PIG	Q48321 sus scrofa
21	27	84.4	585	1 DCE2_RAT	Q05683 rattus norv
22	27	84.4	593	1 DCE1_MOUSE	P48318 mus musculu
23	27	84.4	593	1 DCE1_RAT	P18088 rattus norv
24	27	84.4	594	1 DCE1_FELCA	P14748 felis silve
25	27	84.4	594	1 DCE1_HUMAN	Q09259 homo sapien
26	27	84.4	594	1 DCE1_PIG	P48319 sus scrofa
27	27	84.4	833	1 DPOL_THERFI	O52225 thermus fil
28	27	84.4	947	1 DPOL_RSVI	O70736 red sea bre
29	26	81.2	115	1 Y068_METUA	Q46392 methanococc
30	26	81.2	167	1 RL9_CHLMU	Q46398 chlamydia m
31	26	81.2	167	1 RL9_CHLTR	O84809 chlamydia t
32	26	81.2	229	1 BCLX_CHICK	O07816 gallus gall
33	26	81.2	294	1 RL5_STYCL	Q26481 styela clau

34	26	81.2	297	1 RL5_HELAN	O65353 helianthus
35	26	81.2	310	1 YARE_RHISN	P55638 rhizobium s
36	26	81.2	401	1 YR07_CAEEL	O09424 caenorhabdi
37	26	81.2	417	1 THIK_YEAST	P27796 saccharomyc
38	26	81.2	430	1 NCAP_RABVA	P15197 rabies viru
39	26	81.2	450	1 NCAP_RABVF	O08314 rabies viru
40	26	81.2	450	1 NCAP_RABVU	O09110 rabies viru
41	26	81.2	499	1 MVIN_AQUAE	O67658 aquifex aeo
42	26	81.2	654	1 RPC3_YEAST	P32349 saccharomyc
43	26	81.2	659	1 POL_CERY	P05400 carnation e
44	26	81.2	666	1 POL_FMYD	P09523 figwort mos
45	26	81.2	674	1 POL_CAMVD	P03556 cauliflower

## ALIGNMENTS

```

RESULT 1
ID      NTCL_HUMAN      STANDARD:      PRT: 2444 AA.
AC      P46531:
DT      01-NOV-1995 (rel. 32, last sequence update)
DT      01-NOV-1995 (rel. 32, last sequence update)
DT      01-FEB-1996 (rel. 33, last annotation update)
DE      NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLATION-
DE      ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
GN      NOTCH1 OR TAN1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      MEDLINE=91347367; PubMed=1831692;
RX      Eilisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
RA      Smith S.D., Sklar J.;
RT      "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT      chromosomal translocations in T lymphoblastic neoplasms.";
RL      Cell 66:649-661(1991).
CC      - FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
CC      ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC      IN SOME T-CELL NEOPLASMS.
CC      - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC      - TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC      BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC      IS FOUND MAINLY IN LYMPHOID TISSUES.
CC      - SIMILARITY: HIGH WITH OTHER NOTCH-TYPE PROTEINS.
CC      - SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC      - SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC      - SIMILARITY: CONTAINS 5 ANK REPEATS.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
CC      EMBL: M73980; AAA60614.1; -.
CC      HSSP: P00740; IIXA.
CC      MIM: 190198; -.
CC      InterPro: IPR000152; -.
CC      InterPro: IPR0000561; -.
CC      InterPro: IPR000800; -.
CC      InterPro: IPR001861; -.
CC      InterPro: IPR002110; -.
CC      Pfam: PF00008; EGF_36.
CC      Pfam: PF00023; ank_6.
CC      Pfam: PF00065; notch_3.
CC      PROSITE: PS50088; ANK_REPEAT_4.
CC      PROSITE: PS50297; ANK_REPEAT_REGION_1.
CC      PROSITE: PS00010; ASX_HYDROXYL_20.

```

DR PROSITE; PS00022; EGF\_1; 34.  
DR PROSITE; PS01186; EGF\_2; 26.  
DR PROSITE; PS01187; EGF\_CA\_18.  
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;  
KW Transmembrane; Signal; Glycoprotein.  
FT SIGNAL 1 18 POTENTIAL.  
FT CHAIN 19 >2444 EXTRACELLULAR LOCUS NOTCH PROTEIN HOMOLOG 1.  
FT DOMAIN 19 1736 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.  
FT TRANSMEM 1737 1757 POTENTIAL.  
FT DOMAIN 1758 >2444 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 20 58 EGF-LIKE 1.  
FT DOMAIN 59 99 EGF-LIKE 2.  
FT DOMAIN 102 139 EGF-LIKE 3.  
FT DOMAIN 140 176 EGF-LIKE 4.  
FT DOMAIN 178 216 EGF-LIKE 5.  
FT DOMAIN 218 255 EGF-LIKE 6.  
FT DOMAIN 257 293 EGF-LIKE 7.  
FT DOMAIN 295 333 EGF-LIKE 8.  
FT DOMAIN 335 371 EGF-LIKE 9.  
FT DOMAIN 372 410 EGF-LIKE 10.  
FT DOMAIN 412 450 EGF-LIKE 11.  
FT DOMAIN 452 488 EGF-LIKE 12.  
FT DOMAIN 490 526 EGF-LIKE 13.  
FT DOMAIN 528 564 EGF-LIKE 14.  
FT DOMAIN 566 601 EGF-LIKE 15.  
FT DOMAIN 603 639 EGF-LIKE 16.  
FT DOMAIN 641 676 EGF-LIKE 17.  
FT DOMAIN 678 714 EGF-LIKE 18.  
FT DOMAIN 716 751 EGF-LIKE 19.  
FT DOMAIN 753 789 EGF-LIKE 20.  
FT DOMAIN 791 827 EGF-LIKE 21.  
FT DOMAIN 829 868 EGF-LIKE 22.  
FT DOMAIN 870 906 EGF-LIKE 23.  
FT DOMAIN 908 944 EGF-LIKE 24.  
FT DOMAIN 946 982 EGF-LIKE 25.  
FT DOMAIN 984 1020 EGF-LIKE 26.  
FT DOMAIN 1022 1058 EGF-LIKE 27.  
FT DOMAIN 1060 1096 EGF-LIKE 28.  
FT DOMAIN 1098 1144 EGF-LIKE 29.  
FT DOMAIN 1146 1182 EGF-LIKE 30.  
FT DOMAIN 1184 1220 EGF-LIKE 31.  
FT DOMAIN 1222 1266 EGF-LIKE 32.  
FT DOMAIN 1268 1306 EGF-LIKE 33.  
FT DOMAIN 1308 1347 EGF-LIKE 34.  
FT DOMAIN 1349 1385 EGF-LIKE 35.  
FT DOMAIN 1388 1427 EGF-LIKE 36.  
FT DOMAIN 1446 1481 LIN/NOTCH 1.  
FT DOMAIN 1482 1523 LIN/NOTCH 2.  
FT DOMAIN 1524 1563 LIN/NOTCH 3.  
FT DOMAIN 1567 1597 ANK 1.  
FT DOMAIN 1591 1621 ANK 2.  
FT DOMAIN 1625 1655 ANK 3.  
FT DOMAIN 1659 1689 ANK 4.  
FT DOMAIN 1693 1723 ANK 5.  
FT DOMAIN 1727 1757 POLY-VAL.  
FT DOMAIN 1761 1791 POLY-ARG.  
FT DOMAIN 1795 1825 POLY-PRO.  
FT DOMAIN 1829 1859 POLY-ALA.  
FT DOMAIN 1863 1893 POLY-GLU.  
FT DOMAIN 1897 1927 POLY-GLY.  
FT DOMAIN 1931 1961 POLY-GLN.  
FT DOMAIN 1965 1995 POLY-PRO.  
FT DOMAIN 1999 2029 BY SIMILARITY.  
FT DOMAIN 2033 2063 BY SIMILARITY.  
FT DOMAIN 2067 2097 BY SIMILARITY.  
FT DOMAIN 2101 2131 BY SIMILARITY.  
FT DOMAIN 2135 2165 BY SIMILARITY.  
FT DOMAIN 2169 2199 BY SIMILARITY.  
FT DOMAIN 2203 2233 BY SIMILARITY.  
FT DOMAIN 2237 2267 BY SIMILARITY.  
FT DOMAIN 2271 2301 BY SIMILARITY.  
FT DOMAIN 2305 2335 BY SIMILARITY.  
FT DOMAIN 2339 2369 BY SIMILARITY.  
FT DOMAIN 2373 2403 BY SIMILARITY.  
FT DOMAIN 2407 2437 BY SIMILARITY.  
FT DOMAIN 2441 2471 BY SIMILARITY.  
FT DOMAIN 2475 2505 BY SIMILARITY.  
FT DOMAIN 2509 2539 BY SIMILARITY.  
FT DOMAIN 2543 2573 BY SIMILARITY.  
FT DOMAIN 2577 2607 BY SIMILARITY.  
FT DOMAIN 2611 2641 BY SIMILARITY.  
FT DOMAIN 2645 2675 BY SIMILARITY.  
FT DOMAIN 2679 2709 BY SIMILARITY.  
FT DOMAIN 2713 2743 BY SIMILARITY.  
FT DOMAIN 2747 2777 BY SIMILARITY.  
FT DOMAIN 2781 2811 BY SIMILARITY.  
FT DOMAIN 2815 2845 BY SIMILARITY.  
FT DOMAIN 2849 2879 BY SIMILARITY.  
FT DOMAIN 2883 2913 BY SIMILARITY.  
FT DOMAIN 2917 2947 BY SIMILARITY.  
FT DOMAIN 2951 2981 BY SIMILARITY.  
FT DOMAIN 2985 3015 BY SIMILARITY.  
FT DOMAIN 3019 3049 BY SIMILARITY.  
FT DOMAIN 3053 3083 BY SIMILARITY.  
FT DOMAIN 3087 3117 BY SIMILARITY.  
FT DOMAIN 3121 3151 BY SIMILARITY.  
FT DOMAIN 3155 3185 BY SIMILARITY.  
FT DOMAIN 3189 3219 BY SIMILARITY.  
FT DOMAIN 3223 3253 BY SIMILARITY.  
FT DOMAIN 3257 3287 BY SIMILARITY.  
FT DOMAIN 3291 3321 BY SIMILARITY.  
FT DOMAIN 3325 3355 BY SIMILARITY.  
FT DOMAIN 3359 3389 BY SIMILARITY.  
FT DOMAIN 3393 3423 BY SIMILARITY.  
FT DOMAIN 3427 3457 BY SIMILARITY.  
FT DOMAIN 3461 3491 BY SIMILARITY.  
FT DOMAIN 3495 3525 BY SIMILARITY.  
FT DOMAIN 3529 3559 BY SIMILARITY.  
FT DOMAIN 3563 3593 BY SIMILARITY.  
FT DOMAIN 3597 3627 BY SIMILARITY.  
FT DOMAIN 3631 3661 BY SIMILARITY.  
FT DOMAIN 3665 3695 BY SIMILARITY.  
FT DOMAIN 3699 3729 BY SIMILARITY.  
FT DOMAIN 3733 3763 BY SIMILARITY.  
FT DOMAIN 3767 3797 BY SIMILARITY.  
FT DOMAIN 3801 3831 BY SIMILARITY.  
FT DOMAIN 3835 3865 BY SIMILARITY.  
FT DOMAIN 3869 3899 BY SIMILARITY.  
FT DOMAIN 3903 3933 BY SIMILARITY.  
FT DOMAIN 3937 3967 BY SIMILARITY.  
FT DOMAIN 3971 4001 BY SIMILARITY.  
FT DOMAIN 4005 4035 BY SIMILARITY.  
FT DOMAIN 4039 4069 BY SIMILARITY.  
FT DOMAIN 4073 4103 BY SIMILARITY.  
FT DOMAIN 4107 4137 BY SIMILARITY.  
FT DOMAIN 4141 4171 BY SIMILARITY.  
FT DOMAIN 4175 4205 BY SIMILARITY.  
FT DOMAIN 4209 4239 BY SIMILARITY.  
FT DOMAIN 4243 4273 BY SIMILARITY.  
FT DOMAIN 4277 4307 BY SIMILARITY.  
FT DOMAIN 4311 4341 BY SIMILARITY.  
FT DOMAIN 4345 4375 BY SIMILARITY.  
FT DOMAIN 4379 4409 BY SIMILARITY.  
FT DOMAIN 4413 4443 BY SIMILARITY.  
FT DOMAIN 4447 4477 BY SIMILARITY.  
FT DOMAIN 4481 4511 BY SIMILARITY.  
FT DOMAIN 4515 4545 BY SIMILARITY.  
FT DOMAIN 4549 4579 BY SIMILARITY.  
FT DOMAIN 4583 4613 BY SIMILARITY.  
FT DOMAIN 4617 4647 BY SIMILARITY.  
FT DOMAIN 4651 4681 BY SIMILARITY.  
FT DOMAIN 4685 4715 BY SIMILARITY.  
FT DOMAIN 4719 4749 BY SIMILARITY.  
FT DOMAIN 4753 4783 BY SIMILARITY.  
FT DOMAIN 4787 4817 BY SIMILARITY.  
FT DOMAIN 4821 4851 BY SIMILARITY.  
FT DOMAIN 4855 4885 BY SIMILARITY.  
FT DOMAIN 4889 4919 BY SIMILARITY.  
FT DOMAIN 4923 4953 BY SIMILARITY.  
FT DOMAIN 4957 4987 BY SIMILARITY.  
FT DOMAIN 4991 5021 BY SIMILARITY.  
FT DOMAIN 5025 5055 BY SIMILARITY.  
FT DOMAIN 5059 5089 BY SIMILARITY.  
FT DOMAIN 5093 5123 BY SIMILARITY.  
FT DOMAIN 5127 5157 BY SIMILARITY.  
FT DOMAIN 5161 5191 BY SIMILARITY.  
FT DOMAIN 5195 5225 BY SIMILARITY.  
FT DOMAIN 5229 5259 BY SIMILARITY.  
FT DOMAIN 5263 5293 BY SIMILARITY.  
FT DOMAIN 5297 5327 BY SIMILARITY.  
FT DOMAIN 5331 5361 BY SIMILARITY.  
FT DOMAIN 5365 5395 BY SIMILARITY.  
FT DOMAIN 5399 5429 BY SIMILARITY.  
FT DOMAIN 5433 5463 BY SIMILARITY.  
FT DOMAIN 5467 5497 BY SIMILARITY.  
FT DOMAIN 5501 5531 BY SIMILARITY.  
FT DOMAIN 5535 5565 BY SIMILARITY.  
FT DOMAIN 5569 5599 BY SIMILARITY.  
FT DOMAIN 5603 5633 BY SIMILARITY.  
FT DOMAIN 5637 5667 BY SIMILARITY.  
FT DOMAIN 5671 5701 BY SIMILARITY.  
FT DOMAIN 5705 5735 BY SIMILARITY.  
FT DOMAIN 5739 5769 BY SIMILARITY.  
FT DOMAIN 5773 5803 BY SIMILARITY.  
FT DOMAIN 5807 5837 BY SIMILARITY.  
FT DOMAIN 5841 5871 BY SIMILARITY.  
FT DOMAIN 5875 5905 BY SIMILARITY.  
FT DOMAIN 5909 5939 BY SIMILARITY.  
FT DOMAIN 5943 5973 BY SIMILARITY.  
FT DOMAIN 5977 6007 BY SIMILARITY.  
FT DOMAIN 6011 6041 BY SIMILARITY.  
FT DOMAIN 6045 6075 BY SIMILARITY.  
FT DOMAIN 6079 6109 BY SIMILARITY.  
FT DOMAIN 6113 6143 BY SIMILARITY.  
FT DOMAIN 6147 6177 BY SIMILARITY.  
FT DOMAIN 6181 6211 BY SIMILARITY.  
FT DOMAIN 6215 6245 BY SIMILARITY.  
FT DOMAIN 6249 6279 BY SIMILARITY.  
FT DOMAIN 6283 6313 BY SIMILARITY.  
FT DOMAIN 6317 6347 BY SIMILARITY.  
FT DOMAIN 6351 6381 BY SIMILARITY.  
FT DOMAIN 6385 6415 BY SIMILARITY.  
FT DOMAIN 6419 6449 BY SIMILARITY.  
FT DOMAIN 6453 6483 BY SIMILARITY.  
FT DOMAIN 6487 6517 BY SIMILARITY.  
FT DOMAIN 6521 6551 BY SIMILARITY.  
FT DOMAIN 6555 6585 BY SIMILARITY.  
FT DOMAIN 6589 6619 BY SIMILARITY.  
FT DOMAIN 6623 6653 BY SIMILARITY.  
FT DOMAIN 6657 6687 BY SIMILARITY.  
FT DOMAIN 6691 6721 BY SIMILARITY.  
FT DOMAIN 6725 6755 BY SIMILARITY.  
FT DOMAIN 6759 6789 BY SIMILARITY.  
FT DOMAIN 6793 6823 BY SIMILARITY.  
FT DOMAIN 6827 6857 BY SIMILARITY.  
FT DOMAIN 6861 6891 BY SIMILARITY.  
FT DOMAIN 6895 6925 BY SIMILARITY.  
FT DOMAIN 6929 6959 BY SIMILARITY.  
FT DOMAIN 6963 6993 BY SIMILARITY.  
FT DOMAIN 6997 7027 BY SIMILARITY.  
FT DOMAIN 7031 7061 BY SIMILARITY.  
FT DOMAIN 7065 7095 BY SIMILARITY.  
FT DOMAIN 7099 7129 BY SIMILARITY.  
FT DOMAIN 7133 7163 BY SIMILARITY.  
FT DOMAIN 7167 7197 BY SIMILARITY.  
FT DOMAIN 7201 7231 BY SIMILARITY.  
FT DOMAIN 7235 7265 BY SIMILARITY.  
FT DOMAIN 7269 7299 BY SIMILARITY.  
FT DOMAIN 7303 7333 BY SIMILARITY.  
FT DOMAIN 7337 7367 BY SIMILARITY.  
FT DOMAIN 7371 7401 BY SIMILARITY.  
FT DOMAIN 7405 7435 BY SIMILARITY.  
FT DOMAIN 7439 7469 BY SIMILARITY.  
FT DOMAIN 7473 7503 BY SIMILARITY.  
FT DOMAIN 7507 7537 BY SIMILARITY.  
FT DOMAIN 7541 7571 BY SIMILARITY.  
FT DOMAIN 7575 7605 BY SIMILARITY.  
FT DOMAIN 7609 7639 BY SIMILARITY.  
FT DOMAIN 7643 7673 BY SIMILARITY.  
FT DOMAIN 7677 7707 BY SIMILARITY.  
FT DOMAIN 7711 7741 BY SIMILARITY.  
FT DOMAIN 7745 7775 BY SIMILARITY.  
FT DOMAIN 7779 7809 BY SIMILARITY.  
FT DOMAIN 7813 7843 BY SIMILARITY.  
FT DOMAIN 7847 7877 BY SIMILARITY.  
FT DOMAIN 7881 7911 BY SIMILARITY.  
FT DOMAIN 7915 7945 BY SIMILARITY.  
FT DOMAIN 7949 7979 BY SIMILARITY.  
FT DOMAIN 7983 8013 BY SIMILARITY.  
FT DOMAIN 8017 8047 BY SIMILARITY.  
FT DOMAIN 8051 8081 BY SIMILARITY.  
FT DOMAIN 8085 8115 BY SIMILARITY.  
FT DOMAIN 8119 8149 BY SIMILARITY.  
FT DOMAIN 8153 8183 BY SIMILARITY.  
FT DOMAIN 8187 8217 BY SIMILARITY.  
FT DOMAIN 8221 8251 BY SIMILARITY.  
FT DOMAIN 8255 8285 BY SIMILARITY.  
FT DOMAIN 8289 8319 BY SIMILARITY.  
FT DOMAIN 8323 8353 BY SIMILARITY.  
FT DOMAIN 8357 8387 BY SIMILARITY.  
FT DOMAIN 8391 8421 BY SIMILARITY.  
FT DOMAIN 8425 8455 BY SIMILARITY.  
FT DOMAIN 8459 8489 BY SIMILARITY.  
FT DOMAIN 8493 8523 BY SIMILARITY.  
FT DOMAIN 8527 8557 BY SIMILARITY.  
FT DOMAIN 8561 8591 BY SIMILARITY.  
FT DOMAIN 8595 8625 BY SIMILARITY.  
FT DOMAIN 8629 8659 BY SIMILARITY.  
FT DOMAIN 8663 8693 BY SIMILARITY.  
FT DOMAIN 8697 8727 BY SIMILARITY.  
FT DOMAIN 8731 8761 BY SIMILARITY.  
FT DOMAIN 8765 8795 BY SIMILARITY.  
FT DOMAIN 8799 8829 BY SIMILARITY.  
FT DOMAIN 8833 8863 BY SIMILARITY.  
FT DOMAIN 8867 8897 BY SIMILARITY.  
FT DOMAIN 8901 8931 BY SIMILARITY.  
FT DOMAIN 8935 8965 BY SIMILARITY.  
FT DOMAIN 8969 8999 BY SIMILARITY.  
FT DOMAIN 9003 9033 BY SIMILARITY.  
FT DOMAIN 9037 9067 BY SIMILARITY.  
FT DOMAIN 9071 9101 BY SIMILARITY.  
FT DOMAIN 9105 9135 BY SIMILARITY.  
FT DOMAIN 9139 9169 BY SIMILARITY.  
FT DOMAIN 9173 9203 BY SIMILARITY.  
FT DOMAIN 9207 9237 BY SIMILARITY.  
FT DOMAIN 9241 9271 BY SIMILARITY.  
FT DOMAIN 9275 9305 BY SIMILARITY.  
FT DOMAIN 9309 9339 BY SIMILARITY.  
FT DOMAIN 9343 9373 BY SIMILARITY.  
FT DOMAIN 9377 9407 BY SIMILARITY.  
FT DOMAIN 9411 9441 BY SIMILARITY.  
FT DOMAIN 9445 9475 BY SIMILARITY.  
FT DOMAIN 9479 9509 BY SIMILARITY.  
FT DOMAIN 9513 9543 BY SIMILARITY.  
FT DOMAIN 9547 9577 BY SIMILARITY.  
FT DOMAIN 9581 9611 BY SIMILARITY.  
FT DOMAIN 9615 9645 BY SIMILARITY.  
FT DOMAIN 9649 9679 BY SIMILARITY.  
FT DOMAIN 9683 9713 BY SIMILARITY.  
FT DOMAIN 9717 9747 BY SIMILARITY.  
FT DOMAIN 9751 9781 BY SIMILARITY.  
FT DOMAIN 9785 9815 BY SIMILARITY.  
FT DOMAIN 9819 9849 BY SIMILARITY.  
FT DOMAIN 9853 9883 BY SIMILARITY.  
FT DOMAIN 9887 9917 BY SIMILARITY.  
FT DOMAIN 9921 9951 BY SIMILARITY.  
FT DOMAIN 9955 9985 BY SIMILARITY.  
FT DOMAIN 9989 10000 BY SIMILARITY.  
FT DISULFID 31 46 BY SIMILARITY.  
FT DISULFID 48 57 BY SIMILARITY.  
FT DISULFID 63 74 BY SIMILARITY.  
FT DISULFID 68 87 BY SIMILARITY.  
FT DISULFID 89 98 BY SIMILARITY.  
FT DISULFID 106 117 BY SIMILARITY.  
FT DISULFID 111 127 BY SIMILARITY.  
FT DISULFID 129 138 BY SIMILARITY.  
FT DISULFID 144 155 BY SIMILARITY.  
FT DISULFID 149 164 BY SIMILARITY.  
FT DISULFID 166 175 BY SIMILARITY.  
FT DISULFID 182 195 BY SIMILARITY.  
FT DISULFID 189 204 BY SIMILARITY.  
FT DISULFID 206 215 BY SIMILARITY.  
FT DISULFID 222 233 BY SIMILARITY.  
FT DISULFID 227 243 BY SIMILARITY.  
FT DISULFID 245 254 BY SIMILARITY.  
FT DISULFID 261 272 BY SIMILARITY.  
FT DISULFID 266 281 BY SIMILARITY.  
FT DISULFID 283 292 BY SIMILARITY.  
FT DISULFID 299 312 BY SIMILARITY.  
FT DISULFID 306 321 BY SIMILARITY.  
FT DISULFID 323 332 BY SIMILARITY.  
FT DISULFID 339 350 BY SIMILARITY.  
FT DISULFID 344 359 BY SIMILARITY.  
FT DISULFID 361 370 BY SIMILARITY.  
FT DISULFID 376 387 BY SIMILARITY.  
FT DISULFID 381 398 BY SIMILARITY.  
FT DISULFID 400 409 BY SIMILARITY.  
FT DISULFID 416 429 BY SIMILARITY.  
FT DISULFID 423 438 BY SIMILARITY.  
FT DISULFID 440 449 BY SIMILARITY.  
FT DISULFID 456 467 BY SIMILARITY.  
FT DISULFID 461 476 BY SIMILARITY.  
FT DISULFID 478 487 BY SIMILARITY.  
FT DISULFID 494 505 BY SIMILARITY.  
FT DISULFID 499 514 BY SIMILARITY.  
FT DISULFID 516 525 BY SIMILARITY.  
FT DISULFID 532 543 BY SIMILARITY.  
FT DISULFID 537 552 BY SIMILARITY.  
FT DISULFID 554 563 BY SIMILARITY.  
FT DISULFID 570 580 BY SIMILARITY.  
FT DISULFID 575 589 BY SIMILARITY.  
FT DISULFID 591 608 BY SIMILARITY.  
FT DISULFID 607 618 BY SIMILARITY.  
FT DISULFID 612 627 BY SIMILARITY.  
FT DISULFID 629 638 BY SIMILARITY.  
FT DISULFID 645 655 BY SIMILARITY.  
FT DISULFID 650 664 BY SIMILARITY.  
FT DISULFID 666 675 BY SIMILARITY.  
FT DISULFID 682 693 BY SIMILARITY.  
FT DISULFID 687 702 BY SIMILARITY.  
FT DISULFID 704 713 BY SIMILARITY.  
FT DISULFID 720 730 BY SIMILARITY.  
FT DISULFID 725 739 BY SIMILARITY.  
FT DISULFID 741 750 BY SIMILARITY.  
FT DISULFID 757 768 BY SIMILARITY.  
FT DISULFID 762 777 BY SIMILARITY.  
FT DISULFID 779 788 BY SIMILARITY.  
FT DISULFID 795 806 BY SIMILARITY.  
FT DISULFID 800 815 BY SIMILARITY.  
FT DISULFID 817 826 BY SIMILARITY.  
FT DISULFID 825 839 BY SIMILARITY.  
FT DISULFID 841 850 BY SIMILARITY.  
FT DISULFID 857 867 BY SIMILARITY.  
FT DISULFID 874 885 BY SIMILARITY.  
FT DISULFID 879 894 BY SIMILARITY.  
FT DISULFID 896 905 BY SIMILARITY.  
FT DISULFID 912 923 BY SIMILARITY.  
FT DISULFID 917 932 BY SIMILARITY.  
FT DISULFID 934 943 BY SIMILARITY.  
FT DISULFID 948 959 BY SIMILARITY.  
FT DISULFID 963 978 BY SIMILARITY.  
FT DISULFID 983 999 BY SIMILARITY.  
FT DISULFID 1010 1019 BY SIMILARITY.  
FT DISULFID 1026 1037 BY SIMILARITY.  
FT DISULFID 1031 1046 BY SIMILARITY.  
FT DISULFID 1048 1057 BY SIMILARITY.  
FT DISULFID 1064 1075 BY SIMILARITY.  
FT DISULFID 1069 1084 BY SIMILARITY.  
FT DISULFID 1086 1095 BY SIMILARITY.  
FT DISULFID 1102 1123 BY SIMILARITY.  
FT DISULFID 1117 1132 BY SIMILARITY.  
FT DISULFID 1134 1143 BY SIMILARITY.



FT DISULFID 266 281 BY SIMILARITY.  
FT DISULFID 283 292 BY SIMILARITY.  
FT DISULFID 299 312 BY SIMILARITY.  
FT DISULFID 306 321 BY SIMILARITY.  
FT DISULFID 323 332 BY SIMILARITY.  
FT DISULFID 339 350 BY SIMILARITY.  
FT DISULFID 344 359 BY SIMILARITY.  
FT DISULFID 361 370 BY SIMILARITY.  
FT DISULFID 376 387 BY SIMILARITY.  
FT DISULFID 381 398 BY SIMILARITY.  
FT DISULFID 400 429 BY SIMILARITY.  
FT DISULFID 416 438 BY SIMILARITY.  
FT DISULFID 423 449 BY SIMILARITY.  
FT DISULFID 440 467 BY SIMILARITY.  
FT DISULFID 456 476 BY SIMILARITY.  
FT DISULFID 461 487 BY SIMILARITY.  
FT DISULFID 478 505 BY SIMILARITY.  
FT DISULFID 494 514 BY SIMILARITY.  
FT DISULFID 499 525 BY SIMILARITY.  
FT DISULFID 516 543 BY SIMILARITY.  
FT DISULFID 532 552 BY SIMILARITY.  
FT DISULFID 537 563 BY SIMILARITY.  
FT DISULFID 554 580 BY SIMILARITY.  
FT DISULFID 570 589 BY SIMILARITY.  
FT DISULFID 575 600 BY SIMILARITY.  
FT DISULFID 591 618 BY SIMILARITY.  
FT DISULFID 607 627 BY SIMILARITY.  
FT DISULFID 612 638 BY SIMILARITY.  
FT DISULFID 629 655 BY SIMILARITY.  
FT DISULFID 645 664 BY SIMILARITY.  
FT DISULFID 650 675 BY SIMILARITY.  
FT DISULFID 666 693 BY SIMILARITY.  
FT DISULFID 682 702 BY SIMILARITY.  
FT DISULFID 687 713 BY SIMILARITY.  
FT DISULFID 704 730 BY SIMILARITY.  
FT DISULFID 720 739 BY SIMILARITY.  
FT DISULFID 725 750 BY SIMILARITY.  
FT DISULFID 741 768 BY SIMILARITY.  
FT DISULFID 757 777 BY SIMILARITY.  
FT DISULFID 762 788 BY SIMILARITY.  
FT DISULFID 779 806 BY SIMILARITY.  
FT DISULFID 795 815 BY SIMILARITY.  
FT DISULFID 800 826 BY SIMILARITY.  
FT DISULFID 817 833 BY SIMILARITY.  
FT DISULFID 838 855 BY SIMILARITY.  
FT DISULFID 857 866 BY SIMILARITY.  
FT DISULFID 873 884 BY SIMILARITY.  
FT DISULFID 878 893 BY SIMILARITY.  
FT DISULFID 895 904 BY SIMILARITY.  
FT DISULFID 911 922 BY SIMILARITY.  
FT DISULFID 916 931 BY SIMILARITY.  
FT DISULFID 933 942 BY SIMILARITY.  
FT DISULFID 942 958 BY SIMILARITY.  
FT DISULFID 958 992 BY SIMILARITY.  
FT DISULFID 992 1007 BY SIMILARITY.  
FT DISULFID 1009 1018 BY SIMILARITY.  
FT DISULFID 1025 1036 BY SIMILARITY.  
FT DISULFID 1030 1045 BY SIMILARITY.  
FT DISULFID 1047 1056 BY SIMILARITY.  
FT DISULFID 1063 1074 BY SIMILARITY.  
FT DISULFID 1068 1083 BY SIMILARITY.  
FT DISULFID 1085 1094 BY SIMILARITY.  
FT DISULFID 1101 1122 BY SIMILARITY.  
FT DISULFID 1116 1131 BY SIMILARITY.  
FT DISULFID 1133 1142 BY SIMILARITY.  
FT DISULFID 1149 1160 BY SIMILARITY.

Query Match 100.0%; Score 32; DB 1; Length 2531;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0;

1 GVLRSRK 7  
|||||||

Db 1743 GVLRSRK 1749  
RESULT 3  
NTCL\_RAT ID NTCL\_RAT STANDARD; PRT; 2531 AA.  
DT 007008;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.  
CN NOTCH1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
CX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Schwann cell;  
RA MEDLINE=9211383; PubMed=1764995;  
RA Weinmaster G., Roberts V.J., Lemke G.;  
RT "A homolog of Drosophila Notch expressed during mammalian  
development."  
RL development 113:199-205(1991).  
[2]  
RP REVISIONS TO 1652-1653.  
RA Weinmaster G.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DDA databases.  
CC -1- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER  
CC OF TISSUES.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN  
CC DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE  
CC ADULT.  
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.  
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 3 LNK/NOTCH REPEATS.  
CC -1- SIMILARITY: CONTAINS 5 LNK REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: X57405; CAA40667.1; -  
CC HSSP: P00740; 11XA.  
DR InterPro: IPR000152; -  
DR InterPro: IPR000561; -  
DR InterPro: IPR000800; -  
DR InterPro: IPR001438; -  
DR InterPro: IPR001881; -  
DR InterPro: IPR002049; -  
DR InterPro: IPR002110; -  
DR Pfam: PF00008; EGF\_36.  
DR Pfam: PF00023; ank; 6.  
DR Pfam: PF00066; notch; 3.  
DR PRINTS: PR00010; EGFBLDIN.  
DR PRINTS: PR00011; EGFBLMIN.  
DR PRINTS: PS50088; ANK\_REPEAT; 4.  
DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 22.  
DR PROSITE: PS00022; EGF\_1; 35.  
DR PROSITE: PS01186; EGF\_2; 26.  
DR PROSITE: PS01187; EGF\_CA; 21.  
DR Differentiation: Neurogenesis; Repeat; ANK repeat; EGF-like domain;  
KW Transmembrane; Signal; Glycoprotein.  
KM SIGNAL 1  
FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.  
FT DOMAIN 19 1723 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1724 1746 POTENTIAL.



FT DOMAIN 1747 2531 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 20 58 EGF-Like 1.  
FT DOMAIN 59 99 EGF-Like 2.  
FT DOMAIN 102 139 EGF-Like 3.  
FT DOMAIN 140 176 EGF-Like 4.  
FT DOMAIN 178 216 EGF-Like 5.  
FT DOMAIN 218 255 EGF-Like 6.  
FT DOMAIN 257 293 EGF-Like 7.  
FT DOMAIN 295 333 EGF-Like 8.  
FT DOMAIN 335 371 EGF-Like 9.  
FT DOMAIN 372 410 EGF-Like 10.  
FT DOMAIN 412 450 EGF-Like 11.  
FT DOMAIN 452 488 EGF-Like 12.  
FT DOMAIN 490 526 EGF-Like 13.  
FT DOMAIN 528 564 EGF-Like 14.  
FT DOMAIN 566 601 EGF-Like 15.  
FT DOMAIN 603 639 EGF-Like 16.  
FT DOMAIN 641 676 EGF-Like 17.  
FT DOMAIN 678 714 EGF-Like 18.  
FT DOMAIN 716 751 EGF-Like 19.  
FT DOMAIN 753 789 EGF-Like 20.  
FT DOMAIN 791 827 EGF-Like 21.  
FT DOMAIN 829 867 EGF-Like 22.  
FT DOMAIN 869 905 EGF-Like 23.  
FT DOMAIN 907 943 EGF-Like 24.  
FT DOMAIN 945 981 EGF-Like 25.  
FT DOMAIN 1019 1057 EGF-Like 26.  
FT DOMAIN 1021 1059 EGF-Like 27.  
FT DOMAIN 1059 1095 EGF-Like 28.  
FT DOMAIN 1097 1143 EGF-Like 29.  
FT DOMAIN 1145 1181 EGF-Like 30.  
FT DOMAIN 1183 1219 EGF-Like 31.  
FT DOMAIN 1221 1265 EGF-Like 32.  
FT DOMAIN 1267 1305 EGF-Like 33.  
FT DOMAIN 1307 1344 EGF-Like 34.  
FT DOMAIN 1348 1384 EGF-Like 35.  
FT DOMAIN 1387 1426 EGF-Like 36.  
FT DOMAIN 1449 1462 CYS-RICH.  
FT REPEAT 1917 1946 ANK 1.  
FT REPEAT 1950 1980 ANK 2.  
FT REPEAT 1984 2013 ANK 3.  
FT REPEAT 2017 2046 ANK 4.  
FT REPEAT 2050 2079 ANK 5.  
FT DISULFID 24 37 BY SIMILARITY.  
FT DISULFID 31 46 BY SIMILARITY.  
FT DISULFID 48 57 BY SIMILARITY.  
FT DISULFID 63 74 BY SIMILARITY.  
FT DISULFID 68 87 BY SIMILARITY.  
FT DISULFID 89 98 BY SIMILARITY.  
FT DISULFID 106 117 BY SIMILARITY.  
FT DISULFID 111 127 BY SIMILARITY.  
FT DISULFID 129 138 BY SIMILARITY.  
FT DISULFID 144 155 BY SIMILARITY.  
FT DISULFID 149 164 BY SIMILARITY.  
FT DISULFID 166 175 BY SIMILARITY.  
FT DISULFID 182 195 BY SIMILARITY.  
FT DISULFID 189 204 BY SIMILARITY.  
FT DISULFID 206 215 BY SIMILARITY.  
FT DISULFID 222 233 BY SIMILARITY.  
FT DISULFID 227 243 BY SIMILARITY.  
FT DISULFID 245 254 BY SIMILARITY.  
FT DISULFID 261 272 BY SIMILARITY.  
FT DISULFID 266 281 BY SIMILARITY.  
FT DISULFID 283 292 BY SIMILARITY.  
FT DISULFID 299 312 BY SIMILARITY.  
FT DISULFID 306 321 BY SIMILARITY.  
FT DISULFID 323 332 BY SIMILARITY.  
FT DISULFID 339 350 BY SIMILARITY.  
FT DISULFID 344 359 BY SIMILARITY.  
FT DISULFID 361 370 BY SIMILARITY.  
FT DISULFID 376 387 BY SIMILARITY.  
FT DISULFID 381 398 BY SIMILARITY.  
FT DISULFID 400 409 BY SIMILARITY.

FT DISULFID 416 429 BY SIMILARITY.  
FT DISULFID 423 438 BY SIMILARITY.  
FT DISULFID 440 449 BY SIMILARITY.  
FT DISULFID 456 467 BY SIMILARITY.  
FT DISULFID 461 476 BY SIMILARITY.  
FT DISULFID 478 487 BY SIMILARITY.  
FT DISULFID 494 505 BY SIMILARITY.  
FT DISULFID 499 514 BY SIMILARITY.  
FT DISULFID 516 525 BY SIMILARITY.  
FT DISULFID 532 543 BY SIMILARITY.  
FT DISULFID 537 552 BY SIMILARITY.  
FT DISULFID 554 563 BY SIMILARITY.  
FT DISULFID 570 580 BY SIMILARITY.  
FT DISULFID 575 589 BY SIMILARITY.  
FT DISULFID 591 600 BY SIMILARITY.  
FT DISULFID 607 618 BY SIMILARITY.  
FT DISULFID 612 627 BY SIMILARITY.  
FT DISULFID 629 638 BY SIMILARITY.  
FT DISULFID 645 655 BY SIMILARITY.  
FT DISULFID 650 664 BY SIMILARITY.  
FT DISULFID 666 675 BY SIMILARITY.  
FT DISULFID 682 693 BY SIMILARITY.  
FT DISULFID 687 702 BY SIMILARITY.  
FT DISULFID 704 713 BY SIMILARITY.  
FT DISULFID 720 730 BY SIMILARITY.  
FT DISULFID 725 739 BY SIMILARITY.  
FT DISULFID 741 750 BY SIMILARITY.  
FT DISULFID 757 768 BY SIMILARITY.  
FT DISULFID 762 777 BY SIMILARITY.  
FT DISULFID 779 788 BY SIMILARITY.  
FT DISULFID 795 806 BY SIMILARITY.  
FT DISULFID 800 815 BY SIMILARITY.  
FT DISULFID 817 826 BY SIMILARITY.  
FT DISULFID 833 844 BY SIMILARITY.  
FT DISULFID 838 855 BY SIMILARITY.  
FT DISULFID 857 866 BY SIMILARITY.  
FT DISULFID 873 884 BY SIMILARITY.  
FT DISULFID 878 893 BY SIMILARITY.  
FT DISULFID 895 904 BY SIMILARITY.  
FT DISULFID 911 922 BY SIMILARITY.  
FT DISULFID 916 931 BY SIMILARITY.  
FT DISULFID 933 942 BY SIMILARITY.  
FT DISULFID 987 998 BY SIMILARITY.  
FT DISULFID 992 1007 BY SIMILARITY.  
FT DISULFID 1009 1018 BY SIMILARITY.  
FT DISULFID 1025 1036 BY SIMILARITY.  
FT DISULFID 1030 1045 BY SIMILARITY.  
FT DISULFID 1047 1056 BY SIMILARITY.  
FT DISULFID 1063 1074 BY SIMILARITY.  
FT DISULFID 1068 1083 BY SIMILARITY.  
FT DISULFID 1085 1094 BY SIMILARITY.  
FT DISULFID 1101 1112 BY SIMILARITY.  
FT DISULFID 1116 1131 BY SIMILARITY.  
FT DISULFID 1133 1142 BY SIMILARITY.  
FT DISULFID 1149 1160 BY SIMILARITY.  
FT DISULFID 1154 1169 BY SIMILARITY.  
FT DISULFID 1171 1180 BY SIMILARITY.  
FT DISULFID 1187 1198 BY SIMILARITY.  
FT DISULFID 1192 1207 BY SIMILARITY.

Query Match 100.0%; Score 32; DB 1; Length 2531;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
DB 1743 GVLSRK 1749

RESULT 4  
YQUD\_ECOLI  
ID YQUD\_ECOLI STANDARD; PRT; 101 AA.  
AC P42617;

```

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEORETICAL 11.1 KDA PROTEIN IN EXOR-TDCC INTERGENIC REGION.
GN YQJD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
ON NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RC MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RL -1- SIMILARITY: STROMG, TO E.COLI ELAB AND YGAM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18997; AAA57902.1;
DR EMBL; AE000392; AAC76133.1;
DR Ecogene; EG12743; YqjD.
DR Hypothetical protein.
SQ SEQUENCE 101 AA; 11051 MW; C2CBCE2A482BCC14 CRC64;

Query Match          90.6%; Score 29; DB 1; Length 101;
Best Local Similarity 85.7%; Pred. No. 4.1;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSRK 7
   |||||
Db 95 GVLRSRK 101

RESULT 5
ID RK19_PORPU STANDARD; PRT; 123 AA.
AC P51331;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CHLOROPLAST 50S RIBOSOMAL PROTEIN L19.
GN RPL19.
OS Porphyra purpurea.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
ON NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AVONPORT;
RA Reith M.E., Munnolland J.;
RA "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT genome."
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -1- SIMILARITY: BELONGS TO THE L19P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38804; AAC08217.1;
DR DR Mendel; 10317; PORPU:rp119;1.
DR InterPro; IPR001857;
DR Pfam; PF01245; Ribosomal_L19; 1.
DR PRINTS; PR00061; RIBOSOMAL_L19.
DR PROSITE; PS01015; RIBOSOMAL_L19; 1.
DR Ribosomal protein; Chloroplast.
KW SEQUENCE 123 AA; 14165 MW; 647C4C0359A2154 CRC64;

Query Match          90.6%; Score 29; DB 1; Length 123;
Best Local Similarity 85.7%; Pred. No. 5;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSRK 7
   |||||
Db 52 GVLRSRK 58

RESULT 6
ID LDR_ECOLI STANDARD; PRT; 258 AA.
AC P33233;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PUTATIVE L-LACTATE DEHYDROGENASE OPERON REGULATORY PROTEIN.
GN LDR OR LCTR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
ON NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RC MEDLINE-94012541; PubMed-8407843;
RA Dong J.M., Taylor J.S., Latour D.J., Iuchi S., Lin E.C.C.;
RA "Three overlapping lct genes involved in L-lactate utilization by
RT Escherichia coli."
RT J. Bacteriol. 175:6671-6678(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RC MEDLINE-94316500; PubMed-8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RA "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RT Nucleic Acids Res. 22:2576-2586(1994).
CC -1- FUNCTION: MAY BE A REGULATORY PROTEIN FOR THE LCT GENES.
CC -1- INDUCTION: AEROBICALLY BY L-LACTATE.
CC -1- SIMILARITY: BELONGS TO THE GNTM FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L13970; AAA03584.1;
DR EMBL; U00039; AAB18581.1;
DR EMBL; AE000438; AAC76628.1;
DR PIR; B49904; B49904.
DR Ecogene; EG11962; lldr.
DR InterPro; IPR000524;
DR Pfam; PF00392; gntc; 1.
DR PRINTS; PR00035; HTHGNT.
DR PROSITE; PS00043; HTH_GNTR_FAMILY; 1.
DR Transcription regulation; DNA-binding.
KW

```

FT DNA\_BIND 34 53 H-T-H MOTIF (BY SIMILARITY).  
SQ SEQUENCE 258 AA: 29166 MM; 59C46A3B3456079E CRC64;

Query Match 90.6%; Score 29; DB 1; Length 258;  
Best Local Similarity 85.7%; Pred. No. 10;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVLRSK 7  
Db 59 GVLRSK 65

RESULT 7  
EXUT\_BACSU STANDARD; PRT; 422 AA.  
ID EXUT\_BACSU STANDARD; PRT; 422 AA.  
AC 034456;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HEXURONATE TRANSPORTER.  
GN EXUT.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98240225; PubMed=9579062;  
RA RIVOLTA C., SOLDO B., LAZAREVIC V., JORIS B., MANUEL C., KARAMATA D.;  
RT "A 35.7 kb DNA fragment from the Bacillus subtilis chromosome  
RT containing a putative 12.3 kb operon involved in hexuronate catabolism  
RT and a perfectly symmetrical hypothetical catabolite-responsive  
RT element."  
CC Microbiology 144:877-884(1998).  
CC -1- FUNCTION: ALDOHEXURONATE TRANSPORT SYSTEM.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE PHTHALATE PEROXIDASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: AF015825; AAC46332.1; -;  
DR EMBL: Z99110; CAB13093.1; -;  
DR Subtilist: BG13210; exut.  
DR InterPro: IPR001066; -;  
DR Pfam: PF00083; sugar\_tr. 1.  
KW Transmembrane; Transport.  
KW TRANSMEM 9  
FT TRANSMEM 45 65 POTENTIAL.  
FT TRANSMEM 82 102 POTENTIAL.  
FT TRANSMEM 141 161 POTENTIAL.  
FT TRANSMEM 163 183 POTENTIAL.  
FT TRANSMEM 219 239 POTENTIAL.  
FT TRANSMEM 256 276 POTENTIAL.  
FT TRANSMEM 294 314 POTENTIAL.  
FT TRANSMEM 321 341 POTENTIAL.  
FT TRANSMEM 356 376 POTENTIAL.  
FT TRANSMEM 381 401 POTENTIAL.  
SQ SEQUENCE 422 AA: 45313 MM; C2E291AF347F7EDD CRC64;

Query Match 87.5%; Score 28; DB 1; Length 422;  
Best Local Similarity 85.7%; Pred. No. 30;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVLRSK 7

Db 287 GVLRSK 293

RESULT 8  
PHOQ\_SALTY STANDARD; PRT; 451 AA.  
ID PHOQ\_SALTY STANDARD; PRT; 451 AA.  
AC P14147;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE VIRULENCE SENSOR PROTEIN PHOQ (EC 2.7.3.-).  
GN PHOQ.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 10428;  
RX MEDLINE=89296942; PubMed=2544889;  
RA MILLER S.I., KUKRAL A.M., MEKALANOS J.J.;  
RT "A two-component regulatory system (phoP phoQ) controls Salmonella  
RT typhimurium virulence."  
RT Proc. Natl. Acad. Sci. U.S.A. 86:5054-5058(1989).  
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM PHOQ/PHOP  
CC INVOLVED IN THE REGULATION OF THE EXPRESSION OF GENES INVOLVED IN  
CC VIRULENCE AND MACROPHAGE SURVIVAL OF S. TYPHIMURIUM. PHOQ MAY  
CC FUNCTION AS A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT  
CC PHOSPHORYLATES PHOP IN RESPONSE TO ENVIRONMENTAL SIGNALS.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (PROBABLE).  
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE  
CC KINASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: M24424; AAA27189.1; -;  
DR PIR: B12932; VZEBPT.  
DR StyGene: SG10294; phoQ.  
DR InterPro: IPR000658; -;  
DR Pfam: PF00672; DUF5; 1.  
DR Pfam: PF00512; signal; 1.  
KW Sensory transduction; Transferase; Kinase; Phosphorylation;  
KW Transmembrane; Inner membrane; Virulence.  
KW DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 17 44 POTENTIAL.  
FT DOMAIN 45 170 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 171 198 POTENTIAL.  
FT DOMAIN 199 451 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 236 451 TRANSMITTER DOMAIN (POTENTIAL).  
FT MOD.RES 259 259 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
SQ SEQUENCE 451 AA: 51586 MM; CE0D1E4F7BB43194 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 451;  
Best Local Similarity 85.7%; Pred. No. 32;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVLRSK 7  
Db 313 GVLRSK 319

RESULT 9  
YDS9\_SCHPO

ID YD59\_SCHPO STANDARD; PRT; 786 AA.  
 AC Q10313;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE HYPOCOTYL PROTEIN G6C3.09C IN CHROMOSOME I (FRAGMENT).  
 GN SPAC6C3.09C OR SPAC17G8.01C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Devlin K., Churche C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: STRONG, TO FUNGAL TRNA LIGASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: 269731; CAA93622.1; -  
 DR EMBL: 269795; CAA93684.1; -  
 KW HYPOCOTYL PROTEIN; tRNA processing; Ligase.  
 FT NON-TER  
 SQ SEQUENCE 786 AA; 90211 MW; 51D8F9E0148B91 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 786;  
 Best Local Similarity 85.7%; Pred. No. 57;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVLSRK 7  
 111111  
 Db 624 GVLDRK 630

RESULT 10  
 ID VP2\_AHSV4 STANDARD; PRT; 1060 AA.  
 AC P32553; 064925; 064910; 064927;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE OUTER CAPSID PROTEIN VP2.  
 GN S2 OR L2.  
 OS African horse sickness virus 4 (AHSV-4) (African horse sickness virus  
 OS (serotype 4)).  
 OC Viruses; dsRNA viruses; Reoviridae; Orbiviruses.  
 OX NCBI\_TaxID=36421;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VACCINE;  
 RX MEDLINE=93033117; PubMed=1329319;  
 RA Iwata H., Yamagawa M., Roy P.;  
 RA "Evolutionary relationships among the gnat-transmitted orbiviruses  
 RT that cause African horse sickness, bluetongue, and epizootic  
 RT hemorrhagic disease as evidenced by their capsid protein sequences.";  
 RT Virology 191:251-261(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VACCINE;  
 RX MEDLINE=94355458; PubMed=8075221;  
 RA Sakamoto K., Mizukoshi N., Akiwatnakorn B., Iwata A., Tsuchiya T.,  
 RA Ueda S., Inagawa H., Sugiyama T., Kamada M., Fukusho A.;  
 RA "The complete sequences of African horse sickness virus serotype 4  
 RA (vaccine strain) RNA segment 2 and 6 which encode outer capsid

protein.";  
 RL J. Vet. Med. Sci. 56:321-327(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PLUM ISLAND;  
 RA Stone-Marchat M.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VACCINE;  
 RA Martinez Torrecuadrada J.L., Ranz A., Diez J., Martinez C., Sanz A.,  
 RA Casal J.I.;  
 RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE VP2 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP5)  
 CC WHICH CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID. IT IS THE  
 CC MAJOR TARGET OF THE HOST IMMUNOGENIC RESPONSE.  
 CC -1- SIMILARITY: BELONGS TO THE REOVIRUS VP2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M94680; AAA42537.1; -  
 DR EMBL: D26570; BAA05619.1; -  
 DR EMBL: U21956; AAA67288.1; -  
 DR EMBL: M90697; AAA42541.1; -  
 DR PIR: A36820; P2XRA4.  
 KW Coat protein.  
 FT CONFLICT 9 11  
 FT CONFLICT 9 30  
 FT CONFLICT 158 158  
 FT CONFLICT 163 163  
 FT CONFLICT 169 169  
 FT CONFLICT 176 176  
 FT CONFLICT 178 178  
 FT CONFLICT 190 190  
 FT CONFLICT 196 196  
 FT CONFLICT 232 236  
 FT CONFLICT 261 261  
 FT CONFLICT 295 295  
 FT CONFLICT 304 305  
 FT CONFLICT 331 331  
 FT CONFLICT 342 342  
 FT CONFLICT 346 346  
 FT CONFLICT 357 357  
 FT CONFLICT 360 360  
 FT CONFLICT 384 384  
 FT CONFLICT 396 396  
 FT CONFLICT 402 402  
 FT CONFLICT 418 418  
 FT CONFLICT 427 427  
 FT CONFLICT 437 437  
 FT CONFLICT 466 466  
 FT CONFLICT 470 470  
 FT CONFLICT 481 482  
 FT CONFLICT 489 489  
 FT CONFLICT 499 499  
 FT CONFLICT 555 555  
 FT CONFLICT 583 583  
 FT CONFLICT 587 587  
 FT CONFLICT 602 602  
 FT CONFLICT 652 652  
 FT CONFLICT 662 662  
 FT CONFLICT 681 681  
 FT CONFLICT 691 691  
 FT CONFLICT 715 715  
 FT CONFLICT 730 730  
 FT CONFLICT 766 769  
 MTN -> LTD (IN REF. 2).  
 M -> L (IN REF. 3).  
 K -> T (IN REF. 3).  
 Y -> F (IN REF. 2).  
 K -> G (IN REF. 2).  
 D -> N (IN REF. 2).  
 V -> I (IN REF. 2).  
 Q -> R (IN REF. 3).  
 I -> A (IN REF. 2).  
 E -> Q (IN REF. 2).  
 PNRGG -> QIRG (IN REF. 4).  
 K -> L (IN REF. 3).  
 A -> S (IN REF. 2).  
 IE -> VD (IN REF. 2).  
 A -> T (IN REF. 2).  
 G -> R (IN REF. 2).  
 M -> K (IN REF. 1).  
 G -> R (IN REF. 2).  
 R -> K (IN REF. 2).  
 G -> N (IN REF. 2).  
 K -> R (IN REF. 2).  
 K -> N (IN REF. 2).  
 T -> N (IN REF. 2).  
 V -> I (IN REF. 2).  
 V -> I (IN REF. 3).  
 I -> M (IN REF. 2).  
 S -> G (IN REF. 2).  
 V -> I (IN REF. 2 AND 3).  
 AI -> PR (IN REF. 4).  
 T -> I (IN REF. 2).  
 A -> C (IN REF. 3).  
 A -> T (IN REF. 2).  
 G -> S (IN REF. 2).  
 G -> E (IN REF. 2).  
 Y -> N (IN REF. 4).  
 V -> I (IN REF. 2).  
 V -> I (IN REF. 2).  
 L -> P (IN REF. 2).  
 P -> H (IN REF. 3).  
 A -> T (IN REF. 3).  
 L -> I (IN REF. 3).  
 S -> P (IN REF. 2).  
 TEST -> AYSM (IN REF. 2).

```

FT CONFLICT 780 780 K -> E (IN REF. 2 AND 3).
FT CONFLICT 789 801 HYLETLEQRFND -> LIIMLSNVSLTY (IN REF.
FT CONFLICT 810 810 F -> S (IN REF. 3).
FT CONFLICT 813 813 R -> K (IN REF. 2).
FT CONFLICT 813 813 R -> S (IN REF. 3).
FT CONFLICT 816 816 D -> N (IN REF. 2).
FT CONFLICT 827 827 F -> L (IN REF. 3).
FT CONFLICT 831 831 H -> Q (IN REF. 2).
FT CONFLICT 835 835 E -> Q (IN REF. 2).
FT CONFLICT 864 864 Y -> D (IN REF. 2).
FT CONFLICT 878 878 R -> G (IN REF. 2).
FT CONFLICT 935 935 R -> A (IN REF. 4).
FT CONFLICT 980 980 V -> A (IN REF. 2).
SQ SEQUENCE 1060 AA; 124061 MW; 31F53E6EAD7D12B6 CRC64;

```

Query Match 87.5%; Score 28; DB 1; Length 1060;  
 Best Local Similarity 85.7%; Pred. No. 76;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GVLSRK 7  
 Db 583 GVLSRK 589

```

RESULT 11
RL19_STRTR STANDARD: PRT: 115 AA.
AC 034031.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L19.
GN RPLS.
OS Streptococcus thermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1308;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-CNR2 1205;
RA MEDLINE=96048466; PubMed=9387220;
RA Stanley E., Fitzgerald G.F., le Marec C., Fayard B., van Sinderen D.;
RT "Sequence analysis and characterization of phi O1205, a temperate
  bacteriophage infecting Streptococcus thermophilus CNR21205.";
RL Microbiology 143:3417-3429(1997).
CC -1- FUNCTION: THIS PROTEIN IS LOCATED AT THE 30S-50S RIBOSOMAL SUBUNIT
  INTERFACE AND MAY PLAY A ROLE IN THE STRUCTURE AND FUNCTION OF THE
  AMINOACYL-TRNA BINDING SITE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L19P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 088973; AAC01534.1; -.
CC InterPro: IPR001857; -.
CC Pfam: PF01245; Ribosomal_L19; 1.
CC PRINTS: PR00061; RIBOSOMAL_L19.
CC PROSITE: PS01015; RIBOSOMAL_L19; 1.
CC RIBOSOMAL protein.
SQ SEQUENCE 115 AA; 13146 MW; 376CB96A4EBAB6EB CRC64;

```

Query Match 84.4%; Score 27; DB 1; Length 115;  
 Best Local Similarity 71.4%; Pred. No. 14;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GVLSRK 7  
 Db 45 GVLSRK 51

```

RESULT 12
ID ARRL_ECOLI STANDARD: PRT: 117 AA.
AC P15905;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ARSENICAL RESISTANCE OPERON REPRESSOR.
DE ARSR.
GN Escherichia coli.
OC Plasmid R773.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE=90174986; PubMed=2408017;
RA San Francisco M.J.D., Hope C.L., Owolabi J.B., Tise L.S., Rosen B.P.;
RT "Identification of the metalloregulatory element of the
  plasmid-encoded arsenical resistance operon.";
RL Nucleic Acids Res. 18:619-624(1990).
RN 12
RP FUNCTION.
RX MEDLINE=92157859; PubMed=1838573;
RX Wu J., Rosen B.P.;
RT "The Arsr protein is a trans-acting regulatory protein.";
RL Mol. Microbiol. 5:1331-1336(1991).
RN 13
RP METAL-REGULATION.
RX MEDLINE=93107054; PubMed=8416957;
RX Wu J., Rosen B.P.;
RT "Metalloregulated expression of the ars operon.";
RL J. Biol. Chem. 268:52-58(1993).
CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR FOR THE ARS OPERON. ARSR IS
  A TRANS-ACTING REGULATORY PROTEIN WHICH CONTROLS ITS OWN
  EXPRESSION. THE REPRESSIVE EFFECT OF ARSR IS ALLEVIATED BY OXYGENS
  OF +III OXIDATION STATE OF ARSENIC, ANTIMONY, AND BISMUTH, AS WELL
  AS ARSENATE (AS(V)).
CC -1- SUBUNIT: BINDS DNA AS AN HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE ARSR FAMILY OF TRANSCRIPTIONAL
  REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X16045; CAA34168.1; -.
CC PIR: JS0448; BECCAR.
CC HSP: P30340; ISMT.
CC InterPro: IPR001845; -.
CC Pfam: PF01022; HTH_5; 1.
CC PRINTS: PR00778; HTHARSR.
CC PROSITE: PS00846; HTH_ARSR_FAMILY; 1.
CC Plasmid: Arsenical resistance; Transcription regulation; Repressor;
CC DNA-binding.
FT DNA_BIND 33 52 H-T-H MOTIF (POTENTIAL).
SQ SEQUENCE 117 AA; 13198 MW; 1FDD10766E4F886 CRC64;

```

Query Match 84.4%; Score 27; DB 1; Length 117;  
 Best Local Similarity 71.4%; Pred. No. 14;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

DB 58 GILLDRK 64

```

RESULT 13
CBI1_SALTY STANDARD: PRT: 263 AA.
AC 005591:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 15-JUL-1996 (Rel. 36, Last annotation update)
DE PRECORIN-6X REDUCTASE (EC 1.3.1.54).
GN CBI1.
OS Salmonella typhimurium.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2;
RC MEDLINE=93373696; PubMed=8501034;
RA Roth J.R., Lawrence J.G., Rubenfield M., Kieffer-Higgins S.,
RA Church G.M.;
RT "Characterization of the cobalamin (vitamin B12) biosynthetic genes
RT of Salmonella typhimurium."
RL J. Bacteriol. 175:3303-3316(1993).
CC -1- FUNCTION: CATALYZES THE REDUCTION OF THE MACROCYCLE OF PRECORIN-
CC 6X INTO PRECORIN-6Y (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: PRECORIN-6Y + NADP(+) -> PRECORIN-6X + NADPH.
CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L12006; AAA27261.1; -
CC DR StyGene; SG10042; cblJ.
CC KM Oxidoreductase; NADP; Cobalamin biosynthesis.
CC SQ SEQUENCE 263 AA; 28402 MW; 79E8A4FC5967031C CRC64;

Query Match 84.4%; Score 27; DB 1; Length 263;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLISR 6
DB 94 GVLISR 99

RESULT 14.
RM30_YEAST STANDARD: PRT: 281 AA.
AC P36528;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 60S RIBOSOMAL PROTEIN L30, MITOCHONDRIAL PRECURSOR (YML30).
GN MRPL30 OR YNL252C OR N0864.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS288C / FY1679;
RC MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldeiner U., Beinbauer J.D., Fiedler T.A.,
RA Hegemann J.H.;

```

```

RT "Sequence analysis of the 33 kb long region between ORC5 and SUI1
RT from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL Yeast 13:849-860(1997).
RN [2]
RP SEQUENCE OF 20-40.
RX MEDLINE=91285106; PubMed=2060626;
RA Grohmann L., Graack H.-R., Kruff V., Choi T., Goldschmidt-Reisin S.,
RA Kitakawa M.;
RT "Extended N-terminal sequencing of proteins of the large ribosomal
RT subunit from yeast mitochondria."
RL FEBS Lett. 284:51-56(1991).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X96722; CAA65492.1; -
CC DR EMBL; Z71528; CAA96159.1; -
CC DR PIR: S17270; S17270.
CC DR SCD: S0005196; MRPL30.
CC KW Ribosomal protein; Mitochondrion; Transit peptide.
CC TRANSLIT 1 19
CC CHAIN 20 281 60S RIBOSOMAL PROTEIN L30.
CC FT CHAIN 20 281 60S RIBOSOMAL PROTEIN L30.
CC SQ SEQUENCE 281 AA; 32212 MW; BDAF2B6D7CF44DB CRC64;

```

```

Query Match 84.4%; Score 27; DB 1; Length 281;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLISR 6
DB 28 GVLISR 33

RESULT 15
SAL12_CAEEL STANDARD: PRT: 319 AA.
ID SAL12_CAEEL
AC Q20410;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SRA-12 PROTEIN.
GN SRA-12 OR F44F4.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Coles L.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE C.ELEGANS RECEPTOR-LIKE PROTEIN SRA
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Z37092; CAA85460.1; -
CC DR WormPeP: F44F4.7; CE01587.
CC DR InterPro: IPR000344; -

```

DR Pfam: PF02117; Str: 1.  
 KW Transmembrane; Multigene family.  
 FT TRANSMEM 20 40 POTENTIAL.  
 FT TRANSMEM 54 74 POTENTIAL.  
 FT TRANSMEM 99 119 POTENTIAL.  
 FT TRANSMEM 131 151 POTENTIAL.  
 FT TRANSMEM 178 198 POTENTIAL.  
 FT TRANSMEM 227 247 POTENTIAL.  
 FT TRANSMEM 263 283 POTENTIAL.  
 SQ SEQUENCE 319 AA: 36493 MW: EFF167803BD7D25E CRC64;

Query Match 84.48; Score 27; DB 1; Length 319;  
 Best Local Similarity 85.7%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GYLSRK 7  
 DB 244 GYLSMRK 250

Search completed: September 6, 2001, 16:51:10  
 Job time: 814 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compaq Inc.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:53 : Search time 134.15 seconds  
(without alignments)  
3.975 Million cell updates/sec

Title: US-09-603-713-26

Perfect score: 32

Sequence: 1 GVLSRK 7

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR\_68:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	100.0	2531	2 S18188	notch protein homo
2	32	100.0	2531	2 A46019	notch-1 protein -
3	32	100.0	2555	2 A40043	notch protein homo
4	30	93.8	553	2 S28713	hypothetical prote
5	29	90.6	98	2 T11118	NADH dehydrogenase
6	29	90.6	101	2 G65098	hypothetical 11.1
7	29	90.6	101	2 C85971	hypothetical prote
8	29	90.6	123	2 S73352	ribosomal protein
9	29	90.6	258	1 B49504	probable regulator
10	29	90.6	258	2 C86036	transcription regu
11	28	87.5	301	2 T39401	probable ribose me
12	28	87.5	422	1 A69853	hexonate ribosome
13	28	87.5	487	1 VZEBPT	sensor kinase phoQ
14	28	87.5	719	2 T45845	hypothetical prote
15	28	87.5	786	2 T39034	probable tRNA ligase
16	28	87.5	1059	2 S27891	capsid protein VP2
17	28	87.5	1060	1 S2XR44	outer capsid prote
18	27	84.4	62	2 A85822	hypothetical prote
19	27	84.4	117	1 BVBCAR	arsenical resistan
20	27	84.4	185	2 B71982	probable type II D
21	27	84.4	196	2 E75591	hypothetical prote
22	27	84.4	206	2 F84184	glutamate decarbox
23	27	84.4	206	2 I53274	ribosomal protein
24	27	84.4	281	2 S63225	hypothetical prote
25	27	84.4	319	2 T22191	hypothetical prote
26	27	84.4	366	2 H75180	hypothetical prote
27	27	84.4	397	2 F71125	hypothetical prote
28	27	84.4	436	2 T01652	zinc finger protei
29	27	84.4	448	2 S60961	hypothetical prote

30	27	84.4	471	2 S28476	rfbl protein VC024
31	27	84.4	495	2 S71900	RNA-directed DNA p
32	27	84.4	507	2 G72418	sugar ABC transport
33	27	84.4	539	2 T38454	methionine--trna 1
34	27	84.4	585	1 S38533	glutamate decarbox
35	27	84.4	585	1 A41292	glutamate decarbox
36	27	84.4	585	1 JH0423	glutamate decarbox
37	27	84.4	585	2 S61534	glutamate decarbox
38	27	84.4	585	2 JC4064	glutamate decarbox
39	27	84.4	593	1 A41367	glutamate decarbox
40	27	84.4	593	2 S51776	glutamate decarbox
41	27	84.4	593	2 S48135	glutamate decarbox
42	27	84.4	594	1 B41935	glutamate decarbox
43	27	84.4	594	1 A46758	glutamate decarbox
44	27	84.4	594	2 S51775	glutamate decarbox
45	27	84.4	594	2 JC4065	glutamate decarbox

#### ALIGNMENTS

RESULT 1  
S18188  
notch protein homolog - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 13-Aug-1999  
C:Accession: S18188  
R:Weinmaster, G.; Roberts, V.J.; Lemke, G.  
Development 113, 199-205, 1991  
A:Title: A homolog of Drosophila Notch expressed during mammalian development.  
A:Reference number: S18188; M0ID:92111383  
A:Accession: S18188  
A:Molecule type: mRNA  
A:Residues: 1-2531 <EPI>  
A:Cross-references: EMBL:X57405; NID:957634; PTD:957635  
C:Superfamily: unassigned ankryrin repeat proteins; ankryrin repeat homology; EGF homol  
F:987-1018/Domain: EGF homology <EGF>  
F:1025-1056/Domain: EGF homology <EGF>  
F:1233-1264/Domain: EGF homology <EGF>  
F:1917-1949/Domain: ankryrin repeat homology <AN1>  
F:1950-1982/Domain: ankryrin repeat homology <AN2>  
F:1984-2016/Domain: ankryrin repeat homology <AN3>  
F:2017-2049/Domain: ankryrin repeat homology <AN4>  
F:2050-2082/Domain: ankryrin repeat homology <AN5>

Query Match 100.0%; Score 32; DB 2; Length 2531;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVLSRK 7  
DB 1743 GVLSRK 1749

RESULT 2  
A46019  
Notch-1 protein - mouse  
N:Alternate names: notch protein  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Sep-1999  
C:Accession: A46019; S25144  
R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; G  
Genomics 15, 259-264, 1993  
A:Title: Cloning, analysis, and chromosome localization of Notch-1, a mouse homolog  
A:Reference number: A46019; M0ID:93194170  
A:Accession: A46019  
A:Status: not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-2531 <DEL>  
A:Cross-references: GB:Z11886; GB:S47228; NID:9288502; PTD:CA77941.1; PTD:9288503  
A:Note: sequence extracted from NCBI backbone (NCBIP:12718)  
R:Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J

submitted to the EMBL Data Library, April, 1992  
 A:Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, suggest  
 A:Reference number: S25144  
 A:Accession: S25144  
 A:Molecule type: mRNA  
 A:Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <FRA>  
 A:Cross-references: EMBL:211886  
 C:Genetics:  
 A:Gene: notch-1  
 A:Map position: 2  
 A:Note: proximal region of chromosome 2  
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology  
 F:106-138/Domain: EGF homology <EGF1>  
 F:144-175/Domain: EGF homology <EG01>  
 F:222-254/Domain: EGF homology <EGF2>  
 F:261-292/Domain: EGF homology <EG02>  
 F:339-370/Domain: EGF homology <EG03>  
 F:416-449/Domain: EGF homology <EGF3>  
 F:456-487/Domain: EGF homology <EG04>  
 F:494-525/Domain: EGF homology <EG05>  
 F:532-563/Domain: EGF homology <EG06>  
 F:607-638/Domain: EGF homology <EG07>  
 F:682-713/Domain: EGF homology <EG08>  
 F:757-788/Domain: EGF homology <EG09>  
 F:795-826/Domain: EGF homology <EG10>  
 F:873-904/Domain: EGF homology <EG11>  
 F:911-942/Domain: EGF homology <EG12>  
 F:949-980/Domain: EGF homology <EG13>  
 F:987-1018/Domain: EGF homology <EG14>  
 F:1025-1056/Domain: EGF homology <EG15>  
 F:1063-1094/Domain: EGF homology <EG16>  
 F:1149-1180/Domain: EGF homology <EG17>  
 F:1187-1218/Domain: EGF homology <EG18>  
 F:1233-1264/Domain: EGF homology <EGF4>  
 F:1352-1383/Domain: EGF homology <EG19>  
 F:1391-1425/Domain: EGF homology <EGF>  
 F:1917-1948/Domain: ankyrin repeat homology <AN1>  
 F:1949-1981/Domain: ankyrin repeat homology <AN2>  
 F:1983-2015/Domain: ankyrin repeat homology <AN3>  
 F:2016-2048/Domain: ankyrin repeat homology <AN4>  
 F:2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match 100.0%; Score 32; DB 2; Length 2531;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
 |||||  
 Db 1743 GVLSRK 1749

RESULT 3  
 A40043  
 notch protein homolog TAN-1 precursor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Apr-1992 #sequence\_revision 21-Apr-1992 #text\_change 13-Aug-1999  
 C:Accession: A40043  
 R:Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Sklar, Cell 66, 649-661, 1991  
 A:Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosomal  
 A:Reference number: A40043; MUID:91347367  
 A:Accession: A40043  
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
 A:Molecule type: mRNA  
 A:Residues: 1-2555 <ELL>  
 A:Cross-references: GB:M73980  
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology  
 F:261-292/Domain: EGF homology <EGX1>  
 F:494-525/Domain: EGF homology <EGF1>  
 F:487-1018/Domain: EGF homology <EGX2>  
 F:1149-1180/Domain: EGF homology <EGF>  
 F:1187-1218/Domain: EGF homology <EGF3>

F:1233-1264/Domain: EGF homology <EGX3>  
 F:1927-1959/Domain: ankyrin repeat homology <AN1>  
 F:1960-1992/Domain: ankyrin repeat homology <AN2>  
 F:1994-2026/Domain: ankyrin repeat homology <AN3>  
 F:2027-2059/Domain: ankyrin repeat homology <AN4>  
 F:2060-2092/Domain: ankyrin repeat homology <AN5>

Query Match 100.0%; Score 32; DB 2; Length 2555;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
 |||||  
 Db 1753 GVLSRK 1759

RESULT 4  
 S28713  
 hypothetical protein 4 - sugar beet yellows virus  
 C:Species: sugar beet yellows virus, SBV  
 C:Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 08-Oct-1999  
 C:Accession: S28713  
 R:Agranovsky, A.A.; Boyko, V.P.; Karasev, A.V.; Lunina, N.A.; Koonin, E.V.; Dolja, V. J. Gen. Virol. 72, 15-23, 1991  
 A:Title: Nucleotide sequence of the 3'-terminal half of beet yellows closterovirus RN  
 A:Reference number: S28710; MUID:91116305  
 A:Accession: S28713  
 A:Molecule type: DNA  
 A:Residues: 1-553 <AGR>  
 A:Cross-references: EMBL:X53462; NID:g58878; PIDN:CAA37552.1; PID:g58881

Query Match 93.8%; Score 30; DB 2; Length 553;  
 Best Local Similarity 85.7%; Pred. No. 30;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
 |||||  
 Db 50 GVLSRK 56

RESULT 5  
 T1118  
 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4L - Flourometra serratissima mitoc  
 C:Species: mitochondrion Flourometra serratissima  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 07-Dec-1999  
 C:Accession: T1118  
 R:Scouras, A.; Smith, M.J.  
 submitted to the EMBL Data Library, February 1998  
 A:Description: The complete mitochondrial genome of the crinoid Flourometra serratissi  
 A:Reference number: Z17249  
 A:Accession: T1118  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-98 <SCO>  
 A:Cross-references: EMBL:AF049132; NID:g2970420; PID:g2970425; PIDN:AAD05075.1  
 C:Genetics:  
 A:Genome: mitochondrion  
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 4L  
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylatio

Query Match 90.6%; Score 29; DB 2; Length 98;  
 Best Local Similarity 85.7%; Pred. No. 9.2;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
 |||||  
 Db 18 GVLSRK 24

RESULT 6

G65098  
hypothetical 11.1 kD protein in exnr-tdcc intergenic region - Escherichia coli (strain K12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 29-Sep-1999  
C:Accession: G65098  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
S:Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: G65098  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-101 <BLAT>  
A:Cross-references: GB:AE000392; GB:U00096; NID:92367194; PIDN:AACT6133.1; PID:91789485;  
C:Genetics:  
A:Gene: yqjD  
C:Superfamily: conserved hypothetical protein b2672

Query Match 90.6%; Score 29; DB 2; Length 101;  
Best Local Similarity 85.7%; Pred. No. 9.4;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
|||||  
DB 95 GVLSRK 101

RESULT 7  
C85971  
hypothetical protein yqjD [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: C85971  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Aller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: C85971  
A:Residues: 1-101 <STO>  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-101 <STO>  
A:Cross-references: GB:AE005174; NID:912517688; PIDN:AA658231.1; GSPDB:GN00145; UMGF:244  
A:Experimental source: strain O157:H7, substrain EDJ933  
C:Genetics:  
A:Gene: yqjD  
C:Superfamily: conserved hypothetical protein b2672

Query Match 90.6%; Score 29; DB 2; Length 101;  
Best Local Similarity 85.7%; Pred. No. 9.4;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
|||||  
DB 95 GVLSRK 101

RESULT 8  
S73252  
ribosomal protein L19, chloroplast - red alga (Porphyra purpurea) chloroplast  
C:Species: chloroplast Porphyra purpurea  
C:Date: 19-Mar-1997 #sequence\_revision 09-May-1997 #text\_change 13-Aug-1999  
C:Accession: S73252  
R:Reith, M.; Munholland, J.  
Plant Mol. Biol. Rep. 13, 333-335, 1995  
A:Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.  
A:Reference number: S73108  
A:Accession: S73252  
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-123 <REI>  
A:Cross-references: EMBL:U38804; NID:91276652; PIDN:AA08217.1; PID:91276797  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1995  
C:Genetics:  
A:Gene: rpl19  
A:Genome: chloroplast  
C:Superfamily: Escherichia coli ribosomal protein L19  
C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 90.6%; Score 29; DB 2; Length 123;  
Best Local Similarity 85.7%; Pred. No. 12;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
|||||  
DB 52 GVLSRK 58

RESULT 9  
B49904  
probable regulatory protein lctr - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: B49904; S47825; F65160  
R:Dong, J.M.; Taylor, J.S.; Latour, D.J.; Iuchi, S.; Lin, E.C.C.  
J. Bacteriol. 175, 6671-6678, 1993  
A:Title: Three overlapping lct genes involved in L-lactate utilization by Escherichia  
A:Reference number: A49904; MUID:94012541  
A:Accession: B49904  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-258 <DON>  
A:Cross-references: GB:L13970; NID:9404692; PIDN:AA03584.1; PID:9404694  
R:Plunkett, G.  
Submitted to the EMBL Data Library, March 1994  
A:Reference number: S47666  
A:Accession: S47825  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-258 <PLU>  
A:Cross-references: EMBL:U00039; NID:9466582; PID:9466742  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617  
A:Accession: F65160  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-258 <BLAT>  
A:Cross-references: GB:AE000438; GB:U00096; NID:92367251; PIDN:AACT6628.1; PID:917900  
A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: lctr  
C:Superfamily: regulatory protein fadr  
C:Keywords: DNA binding; transcription regulation

Query Match 90.6%; Score 29; DB 1; Length 258;  
Best Local Similarity 85.7%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
|||||  
DB 59 GVLSRK 65

RESULT 10  
C66036  
transcription regulator [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
 C:Accession: C86036  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 M.L.; Gregory, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouls, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: C86036  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-258 <STO>  
 A:Cross-references: GB:AE005174; NID:q12518354; PIDN:ANG58751.1; GSPDB:GN00145; UWG:Z50  
 A:Experimental source: strain 0157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: *hlyE*  
 C:Superfamily: regulatory protein *fadR*

Query Match 90.6%; Score 29; DB 2; Length 258;  
 Best Local Similarity 85.7%; Pred. No. 24;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
 |||||  
 Db 59 GVLSRK 65

RESULT 11  
 T39401.  
 Probable ribose methyltransferase - fission yeast (*Schizosaccharomyces pombe*)  
 C:Species: *Schizosaccharomyces pombe*  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T39401  
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.  
 submitted to the EMBL Data Library, February 1999  
 A:Reference number: Z21851  
 A:Accession: T39401  
 A:Status: preliminary; translated from GB/EMBL/DDJ  
 A:Molecule type: DNA  
 A:Residues: 1-301 <NCO>  
 A:Cross-references: EMBL:AL035548; PIDN:CAB37444.1; GSPDB:GN00067; SPDB:SPBC1347.13C  
 A:Experimental source: strain 972h-; cosmid c1347  
 C:Genetics:  
 A:Gene: SPDB:SPBC1347.13C  
 A:Map position: 2  
 A:introns: 96/3; 168/3; 260/1

Query Match 87.5%; Score 28; DB 2; Length 301;  
 Best Local Similarity 71.4%; Pred. No. 50;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
 |||||  
 Db 175 GVLSRK 181

RESULT 12  
 A69853  
 hexuronate transporter homolog *yjmg* - *Bacillus subtilis*  
 C:Species: *Bacillus subtilis*  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
 C:Accession: A69853  
 R:Funst, F.; Ogasawara, N.; Mosser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte  
 C.; Bron, S.; Brouillet, S.; Brusch, G.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Ch  
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 399, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
 tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koelters, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
 Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Toononi, A.; Tosato, V.; Uchiya  
 T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A. *Bacillus subtilis*  
 A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: A69853  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-422 <KUN>  
 A:Cross-references: GB:Z99110; GB:AL009126; NID:q263472; PIDN:CAB13093.1; PID:q26335  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: *yjmg*  
 C:Superfamily: hexuronate transporter

Query Match 87.5%; Score 28; DB 1; Length 422;  
 Best Local Similarity 85.7%; Pred. No. 71;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
 |||||  
 Db 287 GVLSRK 293

RESULT 13  
 VZEBT  
 sensor kinase *phoQ* (EC 2.7.3.-) - *Salmonella typhimurium*  
 N:Alternate names: virulence membrane protein *phoQ*  
 C:Species: *Salmonella typhimurium*  
 C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 11-Sep-1998  
 C:Accession: B32932  
 R:Miller, S.I.; Kukral, A.M.; Mekalanos, J.J.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 5054-5058, 1989  
 A>Title: A two-component regulatory system (*phoP phoQ*) controls *Salmonella typhimurium*  
 A:Reference number: A32932; MUID:8926942  
 A:Accession: B32932  
 A:Molecule type: DNA  
 A:Residues: 1-487 <MIL>  
 C:Genetics:  
 A:Gene: *phoQ*  
 A:Map position: 25 min  
 A:Function: phosphorylates *phoP* in response to environmental signals  
 A:Description: two-component regulatory system, *phoP/phoQ* controls the expression of severa  
 A:Note: two-component regulatory system, sensor histidine kinase homology  
 C:Superfamily: *envZ* protein; sensor histidine kinase homology  
 C:Keywords: autophosphorylation; phosphohistidine; phosphoprotein; phosphotransferase  
 F:17-40/Domain: transmembrane #status predicted <TM1>  
 F:189-216/Domain: transmembrane #status predicted <TM2>  
 F:246-478/Domain: sensor histidine kinase homology <SHK>  
 F:277/Binding site: phosphate (His) (covalent) (by autophosphorylation) #status predi

Query Match 87.5%; Score 28; DB 1; Length 487;  
 Best Local Similarity 85.7%; Pred. No. 82;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSRK 7  
 |||||  
 Db 331 GVLSRK 337

RESULT 14  
 T45845  
 hypothetical protein F2K15.260 - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C>Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 18-Feb-2000  
 C:Accession: T45845  
 R:Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Lemck  
 submitted to the Protein Sequence Database, January 2000  
 A:Reference number: Z23015  
 A:Accession: T45845

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-719 <RIE>  
A:Cross-references: EMBL:AL132956  
A:Experimental source: cultivar Columbia; BAC clone F2K15  
C:Genetics:  
A:Map position: 3  
A:Introns: 39/3; 68/1; 128/2; 156/3; 185/3; 215/2; 238/3; 379/2; 406/3; 464/3; 506/3; 58  
A:Note: F2K15.260  
C:Superfamily: Arabidopsis thaliana hypothetical protein F2K15.260

Query Match 87.5%; Score 28; DB 2; Length 719;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVLRSK 7  
|||||  
DB 702 GVLLQK 708

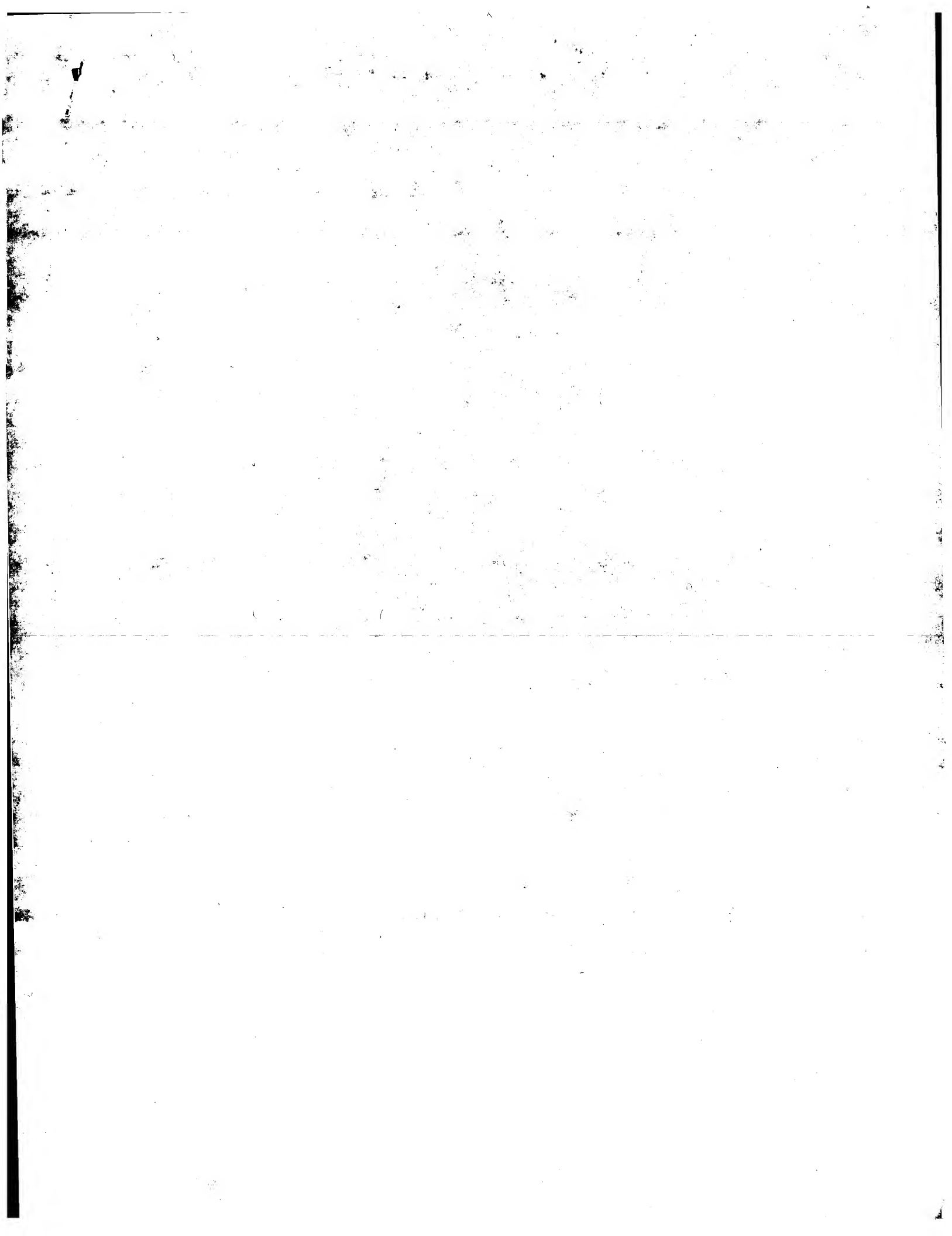
## RESULT 15

T39034  
Probable tRNA ligase - fission yeast (Schizosaccharomyces pombe) (fragment)  
C:Species: Schizosaccharomyces pombe  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 15-Sep-2000  
C:Accession: T39034; T37853  
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z21750  
A:Accession: T39034  
A:Molecule type: DNA  
A:Residues: 3-786 <DEV>  
A:Cross-references: EMBL:Z69731; PIDN:CAA93622.1; GSPDB:GN00066; SPDB:SPAC6C3.10c  
A:Accession: T37853  
A:Molecule type: DNA  
A:Residues: 1-38 <DE2>  
A:Cross-references: EMBL:Z69795; PIDN:CAA93684.1; GSPDB:GN00066; SPDB:SPAC1768.01c

Query Match 87.5%; Score 28; DB 2; Length 786;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVLRSK 7  
|||||  
DB 624 GVLLDRK 630

Search completed: September 6, 2001, 16:45:54  
Job time: 498 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:38 : Search time 113.12 Seconds  
(without alignments)  
1.274 Million cell updates/sec

Title: US-09-603-713-26

Perfect score: 32

Sequence: 1 GVLLSRK 7

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database : Issued Patents, AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfilltest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	1068	1 US-08-537-210A-2	Sequence 2, Appl1
2	32	100.0	1068	4 US-09-113-825-2	Sequence 2, Appl1
3	32	100.0	1078	1 US-08-264-534-32	Sequence 32, Appl1
4	32	100.0	1078	1 US-08-083-590A-11	Sequence 11, Appl1
5	32	100.0	1078	1 US-08-465-500-32	Sequence 32, Appl1
6	32	100.0	1078	2 US-08-346-128-32	Sequence 32, Appl1
7	32	100.0	1078	3 US-08-532-384-11	Sequence 11, Appl1
8	32	100.0	1078	3 US-08-893-828-32	Sequence 32, Appl1
9	32	100.0	2556	1 US-08-185-432-17	Sequence 17, Appl1
10	32	100.0	2556	1 US-08-083-590A-20	Sequence 20, Appl1
11	32	100.0	2556	1 US-08-532-384-20	Sequence 20, Appl1
12	27	84.4	8	4 US-09-056-226-20	Sequence 20, Appl1
13	27	84.4	9	3 US-08-159-339A-950	Sequence 950, App
14	27	84.4	9	3 US-08-159-339A-979	Sequence 979, App
15	27	84.4	10	3 US-08-159-339A-994	Sequence 994, App
16	27	84.4	20	2 US-08-484-530-38	Sequence 38, Appl1
17	27	84.4	20	2 US-08-827-618A-38	Sequence 38, Appl1
18	27	84.4	20	3 US-08-483-952A-38	Sequence 38, Appl1
19	27	84.4	166	2 US-08-592-696-5	Sequence 5, Appl1
20	27	84.4	166	2 US-09-027-536-5	Sequence 5, Appl1
21	27	84.4	166	3 US-09-028-148-5	Sequence 5, Appl1
22	27	84.4	181	2 US-08-308-952-3	Sequence 3, Appl1
23	27	84.4	181	2 US-08-308-952-4	Sequence 4, Appl1
24	27	84.4	181	2 US-08-308-952-5	Sequence 5, Appl1
25	27	84.4	181	4 US-09-124-141-3	Sequence 3, Appl1
26	27	84.4	181	4 US-09-124-141-4	Sequence 4, Appl1
27	27	84.4	181	4 US-09-124-141-5	Sequence 5, Appl1

28	27	84.4	197	4 US-09-124-141-19	Sequence 19, Appl1
29	27	84.4	206	4 US-09-124-141-11	Sequence 11, Appl1
30	27	84.4	436	4 US-09-056-226-2	Sequence 2, Appl1
31	27	84.4	583	6 5475086-4	Patent No. 5475086
32	27	84.4	584	1 US-08-161-290-1	Sequence 1, Appl1
33	27	84.4	584	1 US-08-161-290-2	Sequence 2, Appl1
34	27	84.4	584	2 US-08-450-755-1	Sequence 1, Appl1
35	27	84.4	584	2 US-08-450-755-2	Sequence 2, Appl1
36	27	84.4	585	1 US-08-117-907-2	Sequence 2, Appl1
37	27	84.4	585	1 US-08-485-718-11	Sequence 11, Appl1
38	27	84.4	585	1 US-08-485-718-12	Sequence 12, Appl1
39	27	84.4	585	1 US-08-485-718-13	Sequence 13, Appl1
40	27	84.4	585	2 US-08-484-530-57	Sequence 57, Appl1
41	27	84.4	585	2 US-08-484-530-58	Sequence 58, Appl1
42	27	84.4	585	2 US-08-484-530-59	Sequence 59, Appl1
43	27	84.4	585	2 US-08-494-624-2	Sequence 2, Appl1
44	27	84.4	585	2 US-08-827-618A-57	Sequence 57, Appl1
45	27	84.4	585	2 US-08-827-618A-58	Sequence 58, Appl1

## ALIGNMENTS

RESULT 1  
US-08-537-210A-2  
Sequence 2, Application US/08537210A  
Patent No. 5780300  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Fortini, Mark  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY  
TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036/2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/537,210A  
FILING DATE: 29-SEP-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-027  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1068 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Human NI (TAN-1)  
LOCATION: 1152...2219  
OTHER INFORMATION: Highly conserved ankyrin repeat  
OTHER INFORMATION: region of No. 5780300ch  
US-08-537-210A-2

Query Match 100.0%; Score 32; DB 1; Length 1068;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
|||||||  
Db 603 GVLRSK 609

## RESULT 2

US-09-113-825-2  
Sequence 2, Application US/09113825  
Patent No. 6149902  
GENERAL INFORMATION:  
APPLICANT: Artavans-Tsakonas, Spyridon  
APPLICANT: Fortini, Mark  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY  
TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036/2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/113.825  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/537.210  
FILING DATE: 29-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-027  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1068 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Human N1 (TAN-1)  
LOCATION: 1152...2219  
OTHER INFORMATION: Highly conserved ankyrin repeat  
OTHER INFORMATION: region of No. 6149902ch  
US-09-113-825-2

Query Match 100.0%; Score 32; DB 4; Length 1068;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
|||||||  
Db 603 GVLRSK 609

## RESULT 3

US-08-264-534-32

Sequence 32, Application US/08264534

Patent No. 5648464

GENERAL INFORMATION:

APPLICANT: Artavans-Tsakonas, Spyridon et al.

TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains

TITLE OF INVENTION: In Topolythmic Proteins, And Methods Based Thereon

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/264,534

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/695,189

FILING DATE: 03-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-004

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 8698864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 1078 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-264-534-32

Query Match 100.0%; Score 32; DB 1; Length 1078;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
|||||||  
Db 300 GVLRSK 306

## RESULT 4

US-08-083-590A-11  
Sequence 11, Application US/08083590A  
Patent No. 5786158  
GENERAL INFORMATION:  
APPLICANT: Artavans-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS



SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/083.590A  
FILING DATE: 25-JUN-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MISTOCK, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-083-590A-11

Query Match 100.0%; Score 32; DB 1; Length 1078;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVLSRK 7  
|||||  
DB 300 GVLSRK 306

RESULT 5  
US-08-465-500-32  
Sequence 32, Application US/08465500  
Patent No. 5789195  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Muskavitch, Marc A.T.  
APPLICANT: Fehon, Richard G.  
APPLICANT: Rebay, Ilaria  
APPLICANT: Blaumel, Cristine M.  
APPLICANT: Shepard, Scott B.  
TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS  
TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS, AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,500  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MISTOCK, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-034  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-465-500-32

Query Match 100.0%; Score 32; DB 1; Length 1078;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVLSRK 7  
|||||  
DB 300 GVLSRK 306

RESULT 6  
US-08-346-128-32  
Sequence 32, Application US/08346128  
Patent No. 5856411  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon et al.  
TITLE OF INVENTION: Human No. 5856411ch And Delta, Binding Domains  
TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/346,128  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/879,038  
FILING DATE: 30-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: MISTOCK, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-346-128-32

Query Match 100.0%; Score 32; DB 2; Length 1078;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVLSRK 7  
|||||  
DB 300 GVLSRK 306

RESULT 7  
US-08-532-384-11  
Sequence 11, Application US/08532384  
Patent No. 6083904  
GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And  
TITLE OF INVENTION: Nucleic Acids  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESS: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/532,384  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/083,590  
FILING DATE: 25-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-532-384-11

Query Match 100.0%; Score 32; DB 3; Length 1078;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GYLSRK 7  
1111111  
DB 300 GYLSRK 306

RESULT 8  
US-08-893-828-32  
Sequence 32, Application US/08893828  
Patent No. 6090922  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Muskavitch, Marc A.T.  
APPLICANT: Fehon, Richard G.  
APPLICANT: Rebay, Ilaria  
APPLICANT: Blumuelier, Cristine M.  
TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS  
TITLE OF INVENTION: IN TOPOGRAPHIC PROTEINS, AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESS: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/893,828  
FILING DATE: 11-JUL-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-893-828-32

Query Match 100.0%; Score 32; DB 3; Length 1078;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GYLSRK 7  
1111111  
DB 300 GYLSRK 306

RESULT 9  
US-08-185-432-17  
Sequence 17, Application US/08185432  
Patent No. 5750652  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Bussseau, Isabelle  
APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DETECT PROTEINS, NUCLEIC ACIDS, AND  
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESS: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 256 amino acids

```

; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-17

Query Match          100.0%; Score 32; DB 1; Length 2556;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVLSRK 7
DB 1754 GVLSRK 1760

RESULT 10
US-08-083-590A-20
; Sequence 20, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-083-590A-20

Query Match          100.0%; Score 32; DB 1; Length 2556;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVLSRK 7
DB 1754 GVLSRK 1760

RESULT 11
US-08-532-384-20
; Sequence 20, Application US/08532384
; Patent No. 6083904
; GENERAL INFORMATION:
```

```

; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,384
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,590
; FILING DATE: 25-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-532-384-20

Query Match          100.0%; Score 32; DB 3; Length 2556;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVLSRK 7
DB 1754 GVLSRK 1760

RESULT 12
US-09-056-226-20
; Sequence 20, Application US/09056226B
; Patent No. 6177614
; GENERAL INFORMATION:
; APPLICANT: Colasanti, Joseph J.
; TITLE OF INVENTION: Control of Floral Induction In Plants
; FILE REFERENCE: CSH194-0444
; CURRENT APPLICATION NUMBER: US/09/056,226B
; CURRENT FILING DATE: 1998-04-07
; EARLIER APPLICATION NUMBER: US 09/000,640
; EARLIER FILING DATE: 1997-12-30
; EARLIER APPLICATION NUMBER: US 08/804,104
; EARLIER FILING DATE: 1997-02-20
; EARLIER APPLICATION NUMBER: PCT/US96/03466
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: US 08/406,186
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
```

SEQ ID NO 20  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-056-226-20

Query Match 84.4%; Score 27; DB 4; Length 8;  
Best Local Similarity 71.4%; Pred. No. 1.4e+05;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVLRSK 7  
1:1111  
Db 1 GILFSRK 7

RESULT 13  
US-08-159-339A-950  
Sequence 950, Application US/08159339A  
Patent No. 6037135

GENERAL INFORMATION:  
APPLICANT: Kubo, Ralph T.  
APPLICANT: Grey, Howard M.  
APPLICANT: Sette, Alessandro  
APPLICANT: Celis, Esteban  
TITLE OF INVENTION: HLA Binding peptides and Their  
NUMBER OF SEQUENCES: 1254  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: CA

COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/159,339A  
FILING DATE: 29-NOV-1993

CLASSIFICATION: 424  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/926,666

FILING DATE: 07-AUG-1992  
APPLICATION NUMBER: US 08/027,746

FILING DATE: 05-MAR-1993  
APPLICATION NUMBER: US 08/103,396

FILING DATE: 06-AUG-1993  
ATTORNEY/AGENT INFORMATION:

NAME: Weber, Ellen Lauver

REGISTRATION NUMBER: 32,762

REFERENCE/DOCKET NUMBER: 018623-005030US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

TELEX:  
INFORMATION FOR SEQ ID NO: 950:

SEQUENCE CHARACTERISTICS:

LENGTH: 9 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-159-339A-950

Query Match 84.4%; Score 27; DB 3; Length 9;  
Best Local Similarity 71.4%; Pred. No. 1.4e+05;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVLRSK 7  
1:1111  
Db 1 GILMSRK 7

RESULT 14  
US-08-159-339A-979  
Sequence 979, Application US/08159339A  
Patent No. 6037135

GENERAL INFORMATION:  
APPLICANT: Kubo, Ralph T.  
APPLICANT: Grey, Howard M.  
APPLICANT: Sette, Alessandro  
APPLICANT: Celis, Esteban  
TITLE OF INVENTION: HLA Binding peptides and Their  
NUMBER OF SEQUENCES: 1254  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: CA

COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/159,339A  
FILING DATE: 29-NOV-1993

CLASSIFICATION: 424  
PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 07/926,666  
FILING DATE: 07-AUG-1992

APPLICATION NUMBER: US 08/027,746  
FILING DATE: 05-MAR-1993

APPLICATION NUMBER: US 08/103,396  
FILING DATE: 06-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Ellen Lauver

REGISTRATION NUMBER: 32,762

REFERENCE/DOCKET NUMBER: 018623-005030US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

TELEX:  
INFORMATION FOR SEQ ID NO: 979:

SEQUENCE CHARACTERISTICS:

LENGTH: 9 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-159-339A-979

Query Match 84.4%; Score 27; DB 3; Length 9;  
Best Local Similarity 71.4%; Pred. No. 1.4e+05;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVLRSK 7  
1:1111  
Db 3 GILMSRK 9

RESULT 15

US-08-159-339A-994

Sequence 994, Application US/08159339A

Patent No. 6037135

GENERAL INFORMATION:  
APPLICANT: Kubo, Ralph T.

APPLICANT: Grey, Howard M.  
 APPLICANT: Sette, Alessandro  
 APPLICANT: Celis, Esteban  
 TITLE OF INVENTION: HLA Binding peptides and Their  
 TITLE OF INVENTION: Uses  
 NUMBER OF SEQUENCES: 1254  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend and Crew LLP  
 STREET: Two Embarcadero Center, Eighth Floor  
 CITY: San Francisco  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94111-3634  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/159,339A  
 FILING DATE: 29-NOV-1993  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/926,666  
 FILING DATE: 07-AUG-1992  
 APPLICATION NUMBER: US 08/027,746  
 FILING DATE: 05-MAR-1993  
 APPLICATION NUMBER: US 08/103,396  
 FILING DATE: 06-AUG-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Weber, Ellen Lauver  
 REGISTRATION NUMBER: 32,762  
 REFERENCE/DOCKET NUMBER: 018623-005030US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 576-0200  
 TELEFAX: (415) 576-0300  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 994:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 10 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-159-339A-994

		84.4%	Score 27	DB 3	Length 10
Query Match			Pred. No. 1.5		
Best Local Similarity		71.4%			
Matches	5,	Conservative	2,	Mismatches	0;
				Indels	0;
Gaps					
QY	1 GYLRSRK	7			
	:: :: ::				
Db	4 GILMSRK	10			

Search completed: September 6, 2001, 16:39:39  
Job time: 128 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:31 : Search time 225.25 Seconds  
(without alignments)  
1.884 Million cell updates/sec

Title: US-09-603-713-26

Perfect score: 32

Sequence: 1 GVLSRK 7

## Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

A\_Geneseq\_0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	100.0	7	22	AAB66584
2	32	100.0	7	22	AAB61346
3	32	100.0	10	22	AAB66588
4	32	100.0	10	22	AAB61350
5	32	100.0	1078	13	AAR28963
6	29	90.6	101	21	AAB15976
7	28	87.5	91	22	AA49518
8	28	87.5	172	20	AA59974
9	28	87.5	236	22	AAB8457
10	28	87.5	248	22	AA49503
11	28	87.5	1072	21	AA41664

12	28	87.5	1079	21	AA41663
13	28	87.5	1132	21	AA41662
14	27	84.4	14	16	AAR76655
15	27	84.4	14	18	AA418863
16	27	84.4	20	16	AAR72285
17	27	84.4	20	21	AA59556
18	27	84.4	111	20	AA427358
19	27	84.4	129	21	AA440265
20	27	84.4	166	19	AA474718
21	27	84.4	166	20	AA533657
22	27	84.4	166	21	AA403073
23	27	84.4	181	13	AAR27218
24	27	84.4	181	13	AAR27219
25	27	84.4	186	18	AA455530
26	27	84.4	193	21	AA428206
27	27	84.4	194	18	AA455381
28	27	84.4	285	21	AA410425
29	27	84.4	341	15	AA459522
30	27	84.4	341	15	AA459523
31	27	84.4	343	21	AA410424
32	27	84.4	353	21	AA410423
33	27	84.4	436	20	AA432114
34	27	84.4	438	19	AA459836
35	27	84.4	503	13	AAR23645
36	27	84.4	540	15	AA459520
37	27	84.4	540	15	AA459521
38	27	84.4	544	15	AA459524
39	27	84.4	544	15	AA459525
40	27	84.4	554	15	AA459518
41	27	84.4	554	15	AA459519
42	27	84.4	584	19	AA434519
43	27	84.4	584	19	AA434520
44	27	84.4	584	20	AA486017
45	27	84.4	584	20	AA486018

## ALIGNMENTS

RESULT 1	AA466584	standard; Peptide: 7 AA.
ID	AA466584	
AC	AA466584	
DE	12-APR-2001	(first entry)
XX	Synthetic memapsin 2 substrate peptide #3.	
XX	memapsin 2; nootropic; neuroprotective; amyloid precursor protein;	
KW	APP; memapsin 2 inhibitor; Alzheimer's disease.	
OS	Synthetic.	
XX	WO200100665-A2.	
PN	04-JAN-2001.	
PD	27-JUN-2000; 2000WO-US17742.	
XX	28-JUN-1999; 99US-0141363.	
XX	30-NOV-1999; 99US-0168060.	
PR	25-JAN-2000; 2000US-0177836.	
PR	27-JAN-2000; 2000US-0178368.	
XX	08-JUN-2000; 2000US-0210292.	
PA	(OKLA-) OKLAHOMA MEDICAL RES FOUND.	
XX	(UNITI ) UNIV ILLINOIS FOUND.	
PI	Tang JIN, Hong L, Ghosh AK;	
XX	WPI; 2001-137933/14.	
XX		

Arabidopsis thaliana  
Arabidopsis thaliana  
Peptide derived fr  
65 kD Glutamic aci  
Glutamic acid deca  
GAD65 fragment, pe  
Group B Streptococ  
Human ORFX ORF29 p  
Amino acid sequenc  
Chimeric rat GAD65  
Rat GAD65 glutamic  
Brain GAD. Homo s  
Islet GAD. Homo s  
H. pylori ORF hp2p  
Novel human protei  
H. pylori ORF 09ap  
Arabidopsis thaliana  
GAD65 1-244 N-term  
GAD65 1-244 N-term  
Arabidopsis thaliana  
Arabidopsis thaliana  
Maize id protein.  
Maize id protein.  
Human GAD65 gene p  
GAD65 1-45 N-term  
GAD65 1-45 N-term  
GAD65 545-585 C-te  
GAD65 545-585 C-te  
GAD65 1-31 deleted  
GAD65 1-31 deleted  
Human GAD65 protei  
Rat GAD65 protein  
Human GAD65 protei  
Rat GAD65 protein

PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage  
 XX  
 PS Example 4; Page 33; 86pp; English.  
 CC The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 CC  
 SO Sequence 7 AA;

Query Match 100.0%; Score 32; DB 22; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSSRK 7  
 |||||  
 DB 1 gvlssrk 7

RESULT 2

AAB61346  
 ID AAB61346 standard; peptide; 7 AA.  
 XX  
 AC AAB61346;

DT 02-APR-2001 (first entry)

DE Memapsin 2 substrate #5.

XX Memapsin 2; catalyst; Alzheimer's.

OS Synthetic.

PN WO200100663-A2.

PD 04-JAN-2001.

PF 27-JUN-2000; 2000WO-US17661.

PR 28-JUN-1999; 9905-0141363.

PR 30-NOV-1999; 9905-0168060.

PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.

PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX Tang JTN, Lin X, Koelsch G;

XX WPI: 2001-102885/11.

XX Purified recombinant catalytically active memapsin 2, used to screen

XX inhibitors of it, which are used to treat and prevent Alzheimer's

XX disease -

XX Example 4; Page 33; 86pp; English.

XX The present invention relates to a purified recombinant

XX catalytically active memapsin 2. The invention may be used for

XX isolating inhibitors which are used to treat or prevent

XX Alzheimer's disease. The invention may also be used to screen

XX for individuals more genetically prone to develop Alzheimer's

XX disease.

XX  
 SO Sequence 7 AA;

Query Match 100.0%; Score 32; DB 22; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLSSRK 7  
 |||||  
 DB 1 gvlssrk 7

RESULT 3

AAB65588  
 ID AAB65588 standard; Peptide; 10 AA.  
 XX  
 AC AAB65588;

DT 12-APR-2001 (first entry)

DE Synthetic memapsin 2 substrate peptide.

XX memapsin 2; neurotrophic; neuroprotective; amyloid precursor protein;

XX APP; memapsin 2 inhibitor; Alzheimer's disease.

OS Synthetic.

PN WO200100665-A2.

PD 04-JAN-2001.

PF 27-JUN-2000; 2000WO-US17742.

PR 28-JUN-1999; 9905-0141363.

PR 30-NOV-1999; 9905-0168060.

PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.

PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX (UNIT ) UNIV ILLINOIS FOUND.

XX Tang JTN, Hong L, Ghosh AK;

XX WPI: 2001-137933/14.

XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2

XX having 2 catalytic aspartic residues and substrate binding cleft, used

XX to treat Alzheimer's disease by blocking amyloid precursor protein

XX cleavage

XX Example 4; Page 33; 86pp; English.

XX The present sequence is given in a specification relating to an inhibitor

XX of catalytically active memapsin 2. The inhibitor binds to the memapsin 2

XX active site, which is defined by the presence of two catalytic aspartic

XX residues and a substrate binding cleft. The inhibitor is useful for

XX the treatment and diagnosis of Alzheimer's disease. It is useful in

XX screens for individuals with a genetic predisposition to Alzheimer's

XX disease. The inhibitor is useful as a reagent for specifically binding to

XX memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2

XX isolation, purification and characterisation.

XX Sequence 10 AA;

XX Query Match 100.0%; Score 32; DB 22; Length 10;

XX Best Local Similarity 100.0%; Pred. No. 0.31;

XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX OY 1 GVLSSRK 7

XX |||||



DB 4 gvl1srk 10

RESULT 4

ID AAB61350 standard; peptide; 10 AA.

XX AAB61350;

XX 02-APR-2001 (first entry)

DE Synthetic peptide #2.

XX Memapsin 2; catalyst; Alzheimer's.

XX Synthetic.

OS WO200100663-A2.

PN 04-JAN-2001.

XX 27-JUN-2000; 2000WO-US17661.

PE 28-JUN-1999; 99US-0141363.

PR 30-NOV-1999; 99US-0168060.

PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.

PR 08-JUN-2000; 2000US-0210292.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

PA Tang JN, Lin X, Koelsch G;

PI WPI: 2001-102885/11.

DR Purified recombinant catalytically active memapsin 2, used to screen

PT inhibitors of it, which are used to treat and prevent Alzheimer's

PT disease -

PS Disclosure; Page 86; 86pp; English.

XX The present invention relates to a purified recombinant

CC catalytically active memapsin 2. The invention may be used for

CC isolating inhibitors which are used to treat or prevent

CC Alzheimer's disease. The invention may also be used to screen

CC for individuals more genetically prone to develop Alzheimer's

CC disease.

SO Sequence 10 AA;

Query Match 100.0%; Score 32; DB 22; Length 10;

Best Local Similarity 100.0%; Pred. NO. 0.31; Mismatches 0; Gaps 0;

Matches 7; Conservative 0; Indels 0;

OY 1 GVL1SRK 7

DB 4 gvl1srk 10

RESULT 5

ID AAR28963 standard; Protein; 1078 AA.

AC AAR28963;

XX 01-APR-1993 (first entry)

DE Notch hn3k full length clone.

XX Human; Notch; plasmid; cDNA; clone; hn2k; expression library;

KW adhesion domain; Drosophila; epidermal growth factor; EGF;

KW EGF-like; homologous; repeat; ELR; ELR-11; ELR-12; PCR;

KW polymerase chain reaction; primer; cloning vector; Delta; Serrate;

KW neurogenic; topotypic; homotypic; heterotypic; differentiation;

KW quantitation; antibody.

OS Homo sapiens.

XX WO9219734-A.

XX 12-NOV-1992.

XX 01-MAY-1992; 92WO-US03651.

PE 03-MAY-1991; 91US-0695189.

PR 14-NOV-1991; 91US-0791923.

XX (INDV ) UNIV INDIANA FOUND.

PA (UYTA ) UNIV YALE.

PI Aravanis-Tsakonas S, Blaumueller CM, Fehon RG, Muskavitch MAT;

PI Rebay I, Shepard SB;

XX WPI: 1992-398861/48.

DR N-PSDB; ANQ30999.

XX Human Notch and Delta DNA and protein sequences - used for study

PT and manipulation of differentiation processes

PS Claim 132; Fig 23; 23pp; English.

XX The sequence given represent the full length peptide sequence of

CC human Notch encoded by plasmid cDNA clone hn3k. A human expression

CC library was constructed and screening assays were carried out on the

CC basis of positive binding to the adhesion domain of Notch, ie. the

CC region that is homologous to Drosophila epidermal growth factor

CC (EGF)-like (homologous) repeat (EIR)-11 and -12. Alternatively the

CC sequence could be isolated by amplification using polymerase chain

CC reaction (PCR) primers. The isolated gene may be inserted into a

CC cloning vector and expressed. The Notch gene and also the Delta and

CC serrate neurogenic genes are designated "topotypic" genes. The

CC proteins they encode are involved in specific homo- or heterotypic

CC interactions crucial to differentiation. The quantitation of mRNA

CC for human Notch and Delta and adhesive molecules, and study of its

CC expression are possible using the DNA and antibodies raised against

CC the Notch and Delta proteins.

SO Sequence 1078 AA;

Query Match 100.0%; Score 32; DB 13; Length 1078;

Best Local Similarity 100.0%; Pred. NO. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVL1SRK 7

DB 300 gvl1srk 306

RESULT 6

ID AAB15976 standard; Protein; 101 AA.

AC AAB15976;

XX 05-OCT-2000 (first entry)

DE E. coli proliferation associated protein sequence SEQ ID NO:333.

XX Escherichia coli; E. coli; proliferation; inhibition; screening;

KW antimicrobial; bacterial growth; antisense therapy; antibacterial.

XX Escherichia coli.

OS WO200044906-A2.

PN

XX 03-AUG-2000.  
 PD 27-JAN-2000; 2000MO-US02200.  
 XX PF 27-JAN-1999; 99US-0117405.  
 XX PR 27-JAN-1999; 99US-0117405.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;  
 XX PI Yamamoto R, Xu HH;  
 XX DR WPI: 2000-514822/46.  
 DR N-PSDB; AAA65981.  
 XX PT Novel polynucleotides and polypeptides associated with microorganism  
 PT proliferation, used to identify inhibitors of bacterial growth and  
 PT proliferation, for use in antisense therapy -  
 XX  
 PS Claim 11; Page 249-250; 316pp; English.  
 XX  
 CC AAA65809 to AAA65889 and AAA65058 to AAA66138 represent nucleotide  
 CC sequences derived from Escherichia coli which inhibit E. coli  
 CC proliferation. AAA65890 to AAA66035 and AAA65886 to AAA66040 represent  
 CC nucleotide and protein sequences associated with E. coli proliferation.  
 CC AAA66056 and AAA66057 represent primers used for sequencing E. coli  
 CC proliferation inhibiting nucleotide inserts in an example from the  
 CC present invention. Methods from the present invention can be used to  
 CC identify a proliferation- required gene in a microorganism, by contacting  
 CC a microorganism with a proliferation- required gene activity inhibitory  
 CC nucleic acid identified in another organism, and determining if  
 CC inhibition occurs in the second microorganism. The nucleic acid sequences  
 CC identified as being required for bacterial growth and proliferation, can  
 CC be used for antisense therapy for killing bacteria.  
 XX  
 SO Sequence 101 AA;

Query Match 90.6%; Score 29; DB 21; Length 101;  
 Best Local Similarity 85.7%; Pred. No. 20;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 |||||:  
 DB 95 gvlrsr 101

## RESULT 7

AAAB49518  
 ID AAB49518 standard; Protein; 91 AA.

XX AAB49518;

DT 09-MAR-2001 (first entry)

DE Clone HCEIK90 #1 peptide fragment.

XX Human; 4 transmembrane superfamily receptor protein;  
 KW endocrine; cardiovascular; cerebrovascular disease; neural disorder;  
 KW reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;  
 KW bacterial infection; viral; fungal.

XX Homo sapiens.

PN MO200070076-A1.

PD 23-NOV-2000.

PF 18-MAY-2000; 2000MO-US13504.

PR 19-MAY-1999; 99US-0135122.

PR 03-JUN-1999; 99US-0137797.

PR 11-JUN-1999; 99US-0138573.

PR 18-AUG-1999; 99US-0149447.  
 PR 28-JAN-2000; 2000US-0178770.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Ruben SM, Ni J, Fan P, Roschke V, Shi Y, Komatsoulis GA;  
 XX PI Rosen CA;  
 XX DR WPI: 2001-007502/01.  
 XX PT Isolated nucleic acid molecule encoding human soluble 4 transmembrane  
 PT superfamily receptor protein, useful for diagnosing, treating and/or  
 PT preventing disorders e.g. Alzheimer's, cancer and arthritis -  
 XX  
 PS Disclosure; Page 15; 297pp; English.

XX The present invention relates to isolated nucleic acids and proteins  
 CC encoding human soluble 4 transmembrane superfamily receptor protein (see  
 CC AAC90012-C90023 and AAB49502-B49513). The present sequence is a peptide  
 CC fragment of one such protein. The present protein can be used to screen  
 CC for binding partners and molecules which modify its activity. Antibodies  
 CC specific for the present protein can be used to treat and/or prevent  
 CC diseases associated with aberrant expression or activity of the present  
 CC protein e.g. endocrine disorders e.g. Addison's disease,  
 CC (cardio)vascular diseases e.g. arthritis and atherosclerosis,  
 CC Parkinson's disease, reproductive disorders e.g. Alzheimer's and  
 CC psoriasis, renal system disorders e.g. nephritis, (auto)immune system  
 CC disorders e.g. graft vs. host disease, hyperproliferative disorders e.g.  
 CC neoplasms of the pancreas, ocular disorders e.g. glaucoma and infections  
 CC caused by bacteria, viruses and fungi.

XX Sequence 91 AA;

Query Match 87.5%; Score 28; DB 22; Length 91;  
 Best Local Similarity 85.7%; Pred. No. 30;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVLRSK 7  
 |||||:  
 DB 29 gvlrsr 35

## RESULT 8

AAAY59974  
 ID AAY59974 standard; Protein; 172 AA.

XX AAY59974;

DT 31-JAN-2000 (first entry)

DE Human endometrium tumour EST encoded protein 34.

XX Endometrium; human; tumour; cancer; anticancer; cytostatic; ESR:  
 KW treatment; uterine; gene therapy; expressed sequence tag.

XX Homo sapiens.

PN DE19817948-A1.

PD 21-OCT-1999.

PF 17-APR-1998; 98DE-1017948.

PR 17-APR-1998; 98DE-1017948.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR WPI: 1999-591957/51.

DR N-PSDB; AAZ41992.

XX New nucleic acid sequences expressed in uterine cancer tissues, and  
PT derived polypeptides, for treatment of uterine and endometrial cancer  
PT and identification of therapeutic agents  
PS Claim 23; Page 288; 444pp; German.  
XX  
XX This invention describes novel human nucleic acid (cDNA) sequences (A),  
CC that are highly expressed in uterine tumour tissue and which have  
CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
CC are used (i) to identify agents suitable for treatment of uterine or  
CC endometrial cancer; (ii) directly for treating these forms of cancer  
CC (including expression from gene therapy vectors) and (iii) for  
CC generation of specific antibodies. (A) are identified by assembling ESTs  
CC (expressed sequence tags) from a particular tissue type before comparison  
CC of expression patterns. This allows a significantly longer fragment of  
CC the gene to be revealed, so should reduce the number of failures  
CC associated with the fact that ESTs from different libraries may represent  
CC different parts of the same unknown gene, distorting the estimated  
CC frequency of occurrence in a particular tissue. AAY5941-Y60328 represent  
CC protein fragments encoded by the human endometrium tumour cDNA library  
CC derived EST fragments represented in AAY41981-742121.  
CC  
XX  
SQ Sequence 172 AA;  
Query Match 87.5%; Score 28; DB 20; Length 172;  
Best Local Similarity 85.7%; Pred. No. 62;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GVLSSRK 7  
| | | | |  
DB 51 gvlkrrk 57  
RESULT 9  
AAB88457  
ID AAB88457 standard; Protein: 236 AA.  
XX  
AC AAB88457;  
XX  
DT 23-MAY-2001 (first entry)  
XX  
XX Human membrane or secretory protein clone PSEC0247.  
DE  
XX Human; secretory protein; membrane protein; vaccine; gene therapy;  
KW Rheumatoid arthritis; diabetes.  
XX  
OS Homo sapiens.  
XX  
PN EPI067182-A2.  
XX  
PD 10-JAN-2001.  
XX  
PF 07-JUL-2000; 2000EP-0114090.  
XX  
PR 08-JUL-1999; 99TP-0194179.  
PR 11-JAN-2000; 2000JP-018775.  
PR 02-MAY-2000; 2000JP-0183766.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Mishikawa T, Kawal Y, Sugiyama T, Hayashi K;  
DR WPI: 2001-093989/11.  
DR N-PSDB; AAF93884.  
XX  
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in  
PT gene therapy or as candidate target molecules in drug development -  
XX  
XX Claim 1; SEQ ID 282; 609pp + CD ROM; English.  
XX

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916  
CC which encode human secretory or membrane proteins represented by  
CC AAB88317 - AAB88419. Included in the invention are primers  
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the  
CC cDNA sequences of the invention. The invention also includes methods for  
CC the production of antibodies directed against the proteins, and cDNA  
CC sequences, which can be used in vaccines. The polynucleotide sequences  
CC can be used in gene therapy. The polynucleotide sequences and the  
CC proteins they encode may be used in the prevention, treatment and  
CC diagnosis of diseases associated with inappropriate secretory  
CC protein/membrane protein expression. The nucleic acids and complementary  
CC sequences may also be used as DNA probes in diagnostic assays  
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the  
CC presence of similar nucleic acid sequences in samples. They may also be  
CC used to study the expression and function of secretory proteins/membrane  
CC polypeptides and their role in metabolism. The polypeptides may be used  
CC as antigens in the production of antibodies against them and in assays to  
CC identify modulators (agonists and antagonists) of expression and  
CC activity. The antibodies and antagonists may also be used as therapeutic  
CC agents to down regulate expression and activity. The antibodies may also  
CC be used as diagnostic agents for detecting the presence of the  
CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay  
CC (ELISA). Examples of diseases which may be treated include rheumatoid  
CC arthritis and diabetes.  
CC  
XX  
SQ Sequence 236 AA;  
Query Match 87.5%; Score 28; DB 22; Length 236;  
Best Local Similarity 85.7%; Pred. No. 88;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GVLSSRK 7  
| | | | |  
DB 174 gvlssre 180  
RESULT 10  
AAB49503  
ID AAB49503 standard; Protein: 248 AA.  
XX  
AC AAB49503;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
XX Clone HCEIK90 #1.  
DE  
XX Human; 4 transmembrane superfamily receptor protein;  
KW endocrine; cardiovascular; cerebrovascular disease; neural disorder;  
KW reproductive; skin; renal system; autoimmune; hyperproliferative; ocular;  
KW bacterial infection; viral; fungal.  
XX  
OS Homo sapiens.  
XX  
PN WO20070076-A1.  
XX  
PD 23-NOV-2000.  
XX  
PF 18-MAY-2000; 2000WO-US13504.  
XX  
PR 19-MAY-1999; 99US-0135122.  
PR 03-JUN-1999; 99US-0137797.  
PR 11-JUN-1999; 99US-0138573.  
PR 18-AUG-1999; 99US-0149447.  
PR 28-JAN-2000; 2000US-0178770.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Ni J, Fan P, Roschke V, Shi Y, Komatsoulis GA;  
PI Rosen CA;  
XX  
XX WPI: 2001-007502/01.  
DR N-PSDB; AAC90013.  
DR

XX Isolated nucleic acid molecule encoding human soluble 4 transmembrane  
PT superfamily receptor protein, useful for diagnosing, treating and/or  
PT preventing disorders e.g. Alzheimer's, cancer and arrhythmia -  
XX  
XX  
Claim 11; Pages 279-280; 297pp; English.

CC The present invention relates to isolated nucleic acids and proteins  
CC encoding human soluble 4 transmembrane superfamily receptor protein (see  
CC AC950012, C90023 and AA049502-B49513). The present sequence is one such  
CC protein. The present protein can be used to screen for binding partners  
CC and molecules which modify its activity. Antibodies specific for the  
CC present protein can be used to treat and/or prevent diseases associated  
CC with aberrant expression or activity of the present protein e.g.  
CC endocrine disorders e.g. Addison's disease, (cardio)vascular diseases  
CC e.g. arrhythmia and atherosclerosis, cerebrovascular diseases, neural  
CC disorders e.g. Alzheimer's and Parkinson's disease, reproductive  
CC disorders, skin disorders e.g. psoriasis, renal system disorders e.g.  
CC nephritis, (auto)immune system disorders e.g. graft vs. host disease,  
CC hyperproliferative disorders e.g. neoplasms of the pancreas, ocular  
CC disorders e.g. glaucoma and infections caused by bacteria, viruses and  
CC fungi.

Sequence	248 AA
50	

Query Match	87.5%	Score 28;	DB 22;	length 248;
Best Local	Similarity 85.7%	Pred. NO. 92;		
Matches 6;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0

```
QY      1 GVLLSRK 7
Db      186 gvllsre 192
```

RESULT	11
AAG41664	
ID	AAG41664 standard; Protein; 1072 AA

AC MAG41664;

DT 18-OCT-2000 (first entry)

XX	DE	Arabidopsis thaliana protein fragment	SEQ ID NO:	51867
XX	DE	Arabidopsis thaliana protein fragment	SEQ ID NO:	51867

XX	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter
KM	termination sequence.

OS Arabidopsis thaliana

PN EP1033405-A2

PD 06-SEP-2000

PF 25-FEB-2000; 2000EP-0301439

XX	9905-0121825.
25-FEB-1999.	9905-0123180.
05-MAR-1999.	9905-0123548.
PR 09-MAR-1999.	9905-0125788.
PR 23-MAR-1999.	9905-0125264.
PR 25-MAR-1999.	9905-0126785.
PR 29-MAR-1999.	9905-0127462.
PR 01-APR-1999.	9905-0128234.
PR 06-APR-1999.	9905-0128741.
PR 08-APR-1999.	9905-0129845.
PR 16-APR-1999.	9905-0130077.
PR 19-APR-1999.	9905-0130449.
PR 21-APR-1999.	9905-0130510.
PR 23-APR-1999.	9905-0130891.
PR 23-APR-1999.	9905-0131449.
PR 28-APR-1999.	9905-0131449.
PR 30-APR-1999.	9905-0132048.

[illegible]

PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 12-AUG-1999; 99US-0148319.  
PR 13-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 18-AUG-1999; 99US-0149175.  
PR 20-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156599.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.

PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 87.5%; Score 28; DB 21; Length 1072;  
Best Local Similarity 71.4%; Pred. No. 4.7e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 GVLSRK 7  
I:|||||  
Db 785 gllvstrk 791  
  
RESULT 12  
AAG41663  
ID AAG41663 standard; Protein; 1079 AA.  
XX  
AC AAG41663;  
XX  
DT 18-OCT-2000 (first entry).  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51866.  
XX  
KW Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay: genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129645.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.



PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 87.5%; Score 28; DB 21; Length 1079;  
Best Local Similarity 71.4%; Pred. No. 4.7e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GVLSRK 7  
1:|||||  
DB 792 gllvsrk 798  
  
RESULT 13  
AAC41662  
ID AAC41662 standard; Protein; 1132 AA.  
XX  
XX AAC41662;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51855.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 05-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.

```

PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147492.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159639.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160774.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

```

```

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 87.5%; Score 28; DB 21; Length 1132;
Best Local Similarity 71.4%; Pred. No. 56+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GVLISRK 7
|:|:|
Db 845 gilmstrk 851

```

```

RESULT 14
AAR76655 standard; peptide; 14 AA.
ID AAR76655;
XX AAR76655;

```

```

AC AAR76655;
XX AAR76655;
DT 05-MAR-1996 (first entry)
XX Peptide derived from human glutamic acid decarboxylase 14.
DE diabetes; T-cell subpopulation; detection; antigen production;
XX diagnosis; autoimmune disease.
KW Homo sapiens.
OS DE4418091-AL.
PN 27-JUL-1995.
PD 24-MAY-1994; 94DE-4418091.
PE 04-FEB-1994; 94DE-4403522.
PR 20-JAN-1994; 94DE-4401629.

```

```

XX (ENDL/) ENDL J.
XX (BOEF) BOEHRINGER MANNHEIM GMBH.
PA Albert W, Dormair K, Endl J, Jung G, Meinel E;
PI Stahl P, Schendel D;
XX WPI; 1995-264505/35.
XX

```

```

PT Antigen-specific activated T-lymphocytes and their detection - by
PT interaction with inventive peptide(s) of peptide-MHC complexes;
PT useful in diagnosis of e.g. diabetes and autoimmune diseases
XX

```

```

XX Claim 1; Fig 2; 21pp; German.
XX

```

```

CC AAR76642-62 are derived from human glutamic acid decarboxylase and
CC specifically react with T-cell sub-populations isolated from recently
CC diagnosed Type-I diabetes. Pharmaceutical compns. contg. these
CC peptides and those shown in AAR7571-72, are useful for the diagnosis of
CC a disease or predispositions of immune system diseases, tumours, and
CC autoimmune diseases, including diabetes. The peptides are able to detect
CC specific T-cell subpopulations that are then used for antigen prodn.,
CC e.g. by reinjection.
XX

```

```

XX Sequence 14 AA;
SQ

```

```

Query Match 84.4%; Score 27; DB 16; Length 14;
Best Local Similarity 71.4%; Pred. No. 6.5;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GVLISRK 7
|:|:|
Db 4 gilmstrk 10

```



```

RESULT 15
AAW18863
ID AAW18863 standard; peptide: 14 AA.
XX
AC AAW18863;
XX
DT 05-JAN-1998 (first entry)
XX
DE 65 kD Glutamic acid decarboxylase peptide fragment 14.
XX
KW GAD; 65 kD; human; glutamic acid decarboxylase; autoreactive; diagnosis;
KW insulin-dependent diabetes mellitus; IDDM; cell-mediated disease;
KW prediagnosis; autoimmune; tumour; rheumatoid arthritis;
KW multiple sclerosis.
XX
OS Synthetic.
XX
PN DE19526561-A1.
XX
PD 23-JAN-1997.
XX
PF 20-JUL-1995; 95DE-1026561.
XX
PR 20-JUL-1995; 95DE-1026561.
XX
PA (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI Donle F, Endl J, Ganz M, Jung G, Klentsch-engel R;
PI Pozzilli P, Stahl P;
XX
DR WPI: 1997-088254/09.
XX
PT Skin test for diagnosis of cell-mediated diseases, esp. diabetes -
PT involving intradermal admin. of auto-reactive substances
XX
PS Claim 11; Fig 2; 12pp; German.
XX
CC AAW18842-70 are peptide fragments of the 65 kD human glutamic acid
CC decarboxylase (GAD). The fragments are autoreactive substances used for
CC diagnosis of insulin-dependent diabetes mellitus (IDDM). The diagnosis
CC is determined by using a claimed method for diagnosis of cell-mediated
CC diseases or a prediagnosis to cell-mediated diseases, which is
CC effected by administering an autoreactive substance intradermally and
CC establishing the diagnosis on the basis of the occurrence or lack of a
CC positive reaction at the site of administration. The method is used for
CC diagnosis of autoimmune and tumour diseases, preferably T-cell-mediated
CC diseases such as rheumatoid arthritis, multiple sclerosis and especially
CC IDDM.
XX
XX
Sequence 14 AA:
SO

```

```

Query Match 84.4%; Score 27; DB 18; Length 14;
Best Local Similarity 71.4%; Pred. No. 6.3;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Oy 1 GVLISRK 7
1:|:|:|
Db 4 glimark 10

```

Search completed: September 6, 2001, 16:43:31  
 Job time: 360 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:53 ; Search time 211.42 Seconds  
(without alignments)  
4.002 Million cell updates/sec

Title: US-09-603-713-25  
Perfect score: 32  
Sequence: 1 VCSGYLL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPTRMBL\_16:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phase:\*  
11: sp\_plant:\*  
12: sp\_rhodent:\*  
13: sp\_unclassified:\*  
14: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	802	13	057462
2	31	96.9	966	5	09NKU8
3	30	93.8	227	2	09KTU1
4	30	93.8	510	2	09PC90
5	29	90.6	80	5	023577
6	29	90.6	82	2	09X0T3
7	29	90.6	162	2	09PEP8
8	29	90.6	328	8	09TB46
9	29	90.6	375	10	09SV95
10	29	90.6	380	2	088128
11	29	90.6	421	10	09SYD2
12	29	90.6	442	2	P96400
13	29	90.6	547	2	09X7P8
14	29	90.6	923	2	053890
15	28	87.5	137	2	09X1K2
16	28	87.5	144	3	008439
17	28	87.5	149	2	09RYV8
18	28	87.5	201	2	048808
19	28	87.5	203	2	032344

20	28	87.5	207	2	045491	045491 bacillus st
21	28	87.5	224	2	092537	092537 streptomyc
22	28	87.5	226	5	09NF60	09NF60 drosophila
23	28	87.5	306	5	019660	019660 caenorhabdi
24	28	87.5	319	2	092N76	092N76 streptomyc
25	28	87.5	319	2	0914C5	0914C5 pseudomonas
26	28	87.5	326	2	09RK03	09RK03 streptomyc
27	28	87.5	339	2	09RXF1	09RXF1 deinococcus
28	28	87.5	372	5	023849	023849 dityostell
29	28	87.5	382	2	087918	087918 vibrio chol
30	28	87.5	392	4	09H676	09H676 homo sapien
31	28	87.5	393	2	054031	054031 proplionigen
32	28	87.5	416	2	09PM50	09PM50 campylobact
33	28	87.5	438	2	083469	083469 treponema p
34	28	87.5	439	5	09VZM6	09VZM6 drosophila
35	28	87.5	444	5	021966	021966 caenorhabdi
36	28	87.5	451	2	0912S6	0912S6 pseudomonas
37	28	87.5	462	2	0914P9	0914P9 pseudomonas
38	28	87.5	463	2	09K1H4	09K1H4 neisseria m
39	28	87.5	463	2	09JX25	09JX25 neisseria m
40	28	87.5	487	2	09L3I1	09L3I1 salmonella
41	28	87.5	493	5	017454	017454 caenorhabdi
42	28	87.5	516	5	017455	017455 caenorhabdi
43	28	87.5	545	2	09FB12	09FB12 streptomyc
44	28	87.5	554	2	09XBW6	09XBW6 porphyromon
45	28	87.5	579	5	09U5G9	09U5G9 tetrahymena

## ALIGNMENTS

```
RESULT 1
057462 PRELIMINARY: PRT: 802 AA.
ID 057462:
AC 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE DELTAA.
GN DLA OR DELTAA.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A. Pubmed=9425133;
RX MEDLINE=98165392; Pubmed=9425133;
RA Appel B., Eisen J.S.;
RT "Regulation of neuronal specification in the zebrafish spinal cord by
  Delta function.";
RL Development 125:371-380(1998).
EMBL: AF030031; AAC41249.1; -.
DR HSSP; P00740; LEDM.
DR ZFIN; ZDB-GENE-980526-29; dla.
DR InterPro: IPR000152; -.
DR InterPro: IPR000561; -.
DR InterPro: IPR000742; -.
DR InterPro: IPR001438; -.
DR InterPro: IPR001774; -.
DR InterPro: IPR001881; -.
DR Pfam; PF00008; EGF; 8.
DR Pfam; PF01414; DSL; 1.
DR PRINTS; PR00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_8.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_CA; 1.
DR SMART; SM00001; EGF_1like; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 802 AA; 88941 MW; 0DBA9734585918E5 CRC64;
```

Query Match 100.0%; Score 32; DB 13; Length 802;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 VSGSVLL 7  
 DB 576 VSGSVLL 582

RESULT 2  
 ID 09NKH8 PRELIMINARY; PRT; 966 AA.  
 AC 09NKH8  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 15801.2  
 GN 15801.2  
 OS Leishmania major.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 NCBI\_TaxID=5664;  
 RA SEQUENCE FROM N.A.  
 RC SPATIAL EXPRESSION:  
 RA Myler P.T., Siss E., Hixson G., Kiser P., Rickel E., Hassebrock M.,  
 RA Cawthra J., Marsolini F., Sunkin S., Stuart L.,  
 RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC004145; AAF5396.1; -  
 SO SEQUENCE 966 AA; 101304 MW; AB6BDAEP9AD5012 CRC64;

Query Match 96.9%; Score 31; DB 5; Length 966;  
 Best Local Similarity 95.7%; Pred. No. 5.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 VSGSVLL 7  
 DB 781 VSGSVLL 787

RESULT 3  
 ID 09KTU1 PRELIMINARY; PRT; 227 AA.  
 AC 09KTU1  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 GN BH3268 PROTEIN.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 NCBI\_TaxID=86665;  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP001518; BAB06987.1;  
 SO SEQUENCE 227 AA; 23698 MW; E429009E27B08C2 CRC64;

Query Match 93.8%; Score 30; DB 2; Length 227;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 VSGSVLL 7  
 DB 102 VSGSVLL 108

RESULT 4  
 ID 09PC90  
 AC 09PC90  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 GN XFI1991  
 OS Xylella fastidiosa.  
 OC Xylella  
 OC Xylella; Proteobacteria; gamma subdivision; Xanthomonas group;  
 NCBI\_TaxID=2371;  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RA MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvaranga R., Alves L.M.C., Azeiteiro J.E., Bala G.S., Baptista C.S.,  
 RA Benoit M.R., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,  
 RA Bueno M.R., Camargo A., Camargo L.E.A., Carrer D.M., Carter H.,  
 RA Colauto N.B., Colombo C.A., Costa F., Costa M., Costa M.,  
 RA Coutinho L.L., Cristofari M., Dias-Neto E., Docena C., El-Dorfi H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferris J.R.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,  
 RA Garlier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemp E.L., Kitajima J.P.,  
 RA Klager J.E., Kurume E.E., Laigret F., Lambis M.R., Leite L.C.C.,  
 RA Ramos G.M., Ramos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Marques M.V., Martins E.A.L., Martins M.F., Matsuno C.B.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhami A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA de Oliveira R.R., Pereira G.A., Pereira V.A., Pasquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA da Silva A.C.R., da Silva A.M., da Silva P.R., Sasaki H.E.J.,  
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Melandri J., Setubal J.C.;  
 RT The genome sequence of the plant pathogen xylella fastidiosa.;  
 RT Nature 400:103-107 (2000).  
 DR EMBL: AF004103; AF14697.1; -  
 DR InterPro: IPR00109;  
 DR Pfam: PF00854; PPR2; 2;  
 DR PROSITE: PS01023; PPR2\_2; 1;  
 SO SEQUENCE 510 AA; 56375 MW; 3804023B6ADE29C CRC64;

Query Match 93.8%; Score 30; DB 2; Length 510;  
 Best Local Similarity 71.4%; Pred. No. 4.2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 VSGSVLL 7  
 DB 489 IGSVLL 495

RESULT 5  
 ID 023577 PRELIMINARY; PRT; 80 AA.  
 AC 023577  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 GN KR682.3  
 OS Xylella fastidiosa  
 OC Xylella  
 OC Xylella; Proteobacteria; gamma subdivision; Xanthomonas group;  
 NCBI\_TaxID=2371;  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RA MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvaranga R., Alves L.M.C., Azeiteiro J.E., Bala G.S., Baptista C.S.,  
 RA Benoit M.R., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,  
 RA Bueno M.R., Camargo A., Camargo L.E.A., Carrer D.M., Carter H.,  
 RA Colauto N.B., Colombo C.A., Costa F., Costa M., Costa M.,  
 RA Coutinho L.L., Cristofari M., Dias-Neto E., Docena C., El-Dorfi H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferris J.R.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,  
 RA Garlier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemp E.L., Kitajima J.P.,  
 RA Klager J.E., Kurume E.E., Laigret F., Lambis M.R., Leite L.C.C.,  
 RA Ramos G.M., Ramos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Marques M.V., Martins E.A.L., Martins M.F., Matsuno C.B.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhami A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA de Oliveira R.R., Pereira G.A., Pereira V.A., Pasquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA da Silva A.C.R., da Silva A.M., da Silva P.R., Sasaki H.E.J.,  
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Melandri J., Setubal J.C.;  
 RT The genome sequence of the plant pathogen xylella fastidiosa.;  
 RT Nature 400:103-107 (2000).  
 DR EMBL: AF004103; AF14697.1; -  
 DR InterPro: IPR00109;  
 DR Pfam: PF00854; PPR2; 2;  
 DR PROSITE: PS01023; PPR2\_2; 1;  
 SO SEQUENCE 510 AA; 56375 MW; 3804023B6ADE29C CRC64;

RP SEQUENCE FROM N.A.  
 RX MEDLINE-94150718; PubMed-7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
 RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,  
 RA Watson A., Welnsloock L., Wilkinson-Sproat J., Wohlman P.,  
 RT \*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.\*  
 RL Nature 368:32-38(1994).  
 RP SEQUENCE FROM N.A.  
 RA Du Z., Le T.,  
 RL Submitted (Dec-1995) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.,  
 RL Submitted (Nov-1995) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: U41110; AA82416.1;  
 SO SEQUENCE 80 AA; 9612 MW; 9A1766FD56C75E5C CRC64;

Query Match 90.6%; Score 29; DB 5; Length 80;  
 Best Local Similarity 85.7%; Pred. No. 95;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGVLL 7  
 ||:||||  
 DB 22 VSGVLL 28

RESULT 6  
 ID 09X073 PRELIMINARY; PRT; 82 AA.  
 AC 09X073:  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE CONSERVED HYPOTHETICAL PROTEIN.  
 GN TML206.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogales; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MS8 / DSM 3109;  
 RX MEDLINE-99287316; PubMed-10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,  
 RT \*Evidence for lateral gene transfer between Archaea and Bacteria from  
 RT genome sequence of Thermotoga maritima.\*  
 RL Nature 399:323-329(1999).  
 DR EMBL: AE001778; AAD36281.1;  
 DR TIGR: TML206;  
 SO SEQUENCE 82 AA; 9090 MW; 55D2B15BB48E8FD CRC64;

Query Match 90.6%; Score 29; DB 2; Length 82;  
 Best Local Similarity 85.7%; Pred. No. 97;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGVLL 7  
 ||:||||  
 DB 9 VSGVLL 15

RESULT 7  
 ID 09PEP8 PRELIMINARY; PRT; 162 AA.  
 AC 09PEP8:  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE LIPOPOLYSACCHARIDE SYNTHESIS ENZYME.  
 GN XF0980.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-9A5C;  
 RX MEDLINE-20365717; PubMed-10910347;  
 RA Simpson A.J.G., Reinach F.C., Artuda P., Abreu F.A., Acencio M.,  
 RA Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Klajim J.P.,  
 RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Klajim J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorallo C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Pelixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Queglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshahko M.H.,  
 RA Vallada H., Van Sluys M.A., Verjowski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Melandri J., Setubal J.C.,  
 RT \*The genome sequence of the plant pathogen Xylella fastidiosa.\*  
 RL Nature 406:151-159(2000).  
 DR EMBL: AE003936; AAF83790.1;  
 DR InterPro: IPR001980;  
 DR InterPro: IPR001994;  
 DR Pfam: PF01467; CYLidylyltransf. 1.  
 DR PRINTS: PRO1020; LPSHOSNTHSS.  
 SO SEQUENCE 162 AA; 17568 MW; 519E4052DAE364FD CRC64;

Query Match 90.6%; Score 29; DB 2; Length 162;  
 Best Local Similarity 85.7%; Pred. No. 26+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGVLL 7  
 ||:||||  
 DB 85 VSGVLL 91

RESULT 8  
 ID 09TB46 PRELIMINARY; PRT; 328 AA.  
 AC 09TB46:  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 GN NAD2.  
 DE NADH DEHYDROGENASE SUBUNIT 2.

OS Platyneris dumerilii (Dumeril's clam worm).  
 OC Mitochondrion.  
 CC Eukaryota; Metazoa; Annelida; Polychaeta; Palmeta; Aciculata;  
 CC Phyllocoelidae; Nereididae; Platyneris.  
 OX NCBI\_TaxID=6359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Boore J.L., Brown W.M.;  
 RT "Mitochondrial genomes of Galatheaolina, Helobdella, and Platyneris:  
 RT Sequence and gene arrangement comparisons indicate that Pogonophora is  
 RT not a phylum and that Annelida and Arthropoda are not sister taxa."  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 CC -I- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS  
 CC CHAINS.  
 CC EMBL: AF178678; AF02688.1; -  
 DR InterPro: IPR001750; -  
 DR Pfam: PF00361; oxidored\_g1; 1.  
 KW Mitochondrion; NAD; oxidoreductase.  
 SQ SEQUENCE 328 AA; 36412 MW; FA7FAB40481FE5D CRC64;

Query Match 90.6%; Score 29; DB 8; Length 328;  
 Best Local Similarity 71.4%; Pred. No. 4.2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :||||:  
 DB 64 IGSVLL 70

RESULT 9  
 ID 09SV95 PRELIMINARY; PRT; 375 AA.  
 AC 09SV95;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE HYPOHETICAL 42.3 KDA PROTEIN.  
 GN F24G24.10 OR AT4G10210.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 CC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,  
 RA Mayer K.F.X., Scheller C.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL049488; CAB39774.1; -  
 DR EMBL: AL161516; CAB78144.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 375 AA; 42325 MW; CA6D9AFBD882ECF6 CRC64;

Query Match 90.6%; Score 29; DB 10; Length 375;  
 Best Local Similarity 71.4%; Pred. No. 4.9e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :||||:  
 DB 238 VSGSVLL 244

RESULT 10

088128  
 ID 088128 PRELIMINARY; PRT; 380 AA.  
 AC 088128;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE VPF380.  
 GN VPF380.  
 OS Vibrio parahaemolyticus.  
 CC Plasmid VF12, and plasmid VF33.  
 CC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=670;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98422461; PubMed=9748441;  
 RA Chang B., Taniguchi H., Miyamoto H., Yoshida S.;  
 RT "Filamentous bacteriophages of Vibrio parahaemolyticus as a possible  
 RT clue to genetic transmission."  
 RL J. Bacteriol. 180:5094-5101(1998).  
 DR EMBL: AB012574; BAA33519.1; -  
 DR EMBL: AB012573; BAA33511.1; -  
 KW Plasmid.  
 SQ SEQUENCE 380 AA; 43671 MW; 1A22914738FB01C1 CRC64;

Query Match 90.6%; Score 29; DB 2; Length 380;  
 Best Local Similarity 85.7%; Pred. No. 5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :||||:  
 DB 216 VGTGVL 222

RESULT 11  
 ID 09SYD2 PRELIMINARY; PRT; 421 AA.  
 AC 09SYD2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE F11M15.16 PROTEIN.  
 GN F11M15.16.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 CC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Federpsiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
 RA Altairi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,  
 RA Gonzalez A., Kremetskaia I., Kim C., Lenz C., Li J., Liu S.,  
 RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,  
 RA Walker M., Yu G., Becker J., Theologis A., Davis R.W.;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC006085; AAD30642.1; -  
 SQ SEQUENCE 421 AA; 47495 MW; DD4C7179BEBDB061 CRC64;

Query Match 90.6%; Score 29; DB 10; Length 421;  
 Best Local Similarity 71.4%; Pred. No. 5.5e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :||||:  
 DB 358 IGSVLL 364

RESULT 12  
 ID P96400 PRELIMINARY; PRT; 442 AA.  
 AC P96400;

DT 01-MAY-1997 (TREMblrel. 03, Created)  
 DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
 DE HYPOTHETICAL 47.4 KDA PROTEIN.  
 GN RV0218 OR MFCY08D5.13.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxID=1773;  
 RX NCBI\_TaxID=1773;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE-98295987; PubMed-9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 DR EMBL: Z92669; CAB07013.1; -.  
 DR Tuberculol: RV0218; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 442 AA; 47361 MW; F4FF948CB2DF2CC CRC64;

Query Match 90.6%; Score 29; DB 2; Length 442;  
 Best Local Similarity 85.7%; Pred. No. 5.8e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 DB 253 VGGVLL 259

RESULT 13  
 ID 09X7P8 PRELIMINARY; PRT; 547 AA.  
 AC 09X7P8;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE PUTATIVE ABC TRANSPORTER.  
 GN SC5F2A.25.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 NCBI\_TaxID=1902;  
 RX NCBI\_TaxID=1902;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Oliver K., Harris D.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE-97000351; PubMed-8843436;  
 RA Redenbach M., Kleser H.M., Denapalae D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RL M01. Microbiol. 21:77-96(1996).

DR EMBL: AL049587; CAB40692.1; -.  
 DR InterPro: IPR001687; -.  
 DR InterPro: IPR003439; -.  
 DR InterPro: IPR003593; -.  
 DR Pfam: PF00005; ABC\_tran; 2.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 DR SMART: SM00382; AAA; 1.  
 SQ SEQUENCE 547 AA; 59276 MW; A7527E393408DCC0 CRC64;

Query Match 90.6%; Score 29; DB 2; Length 547;  
 Best Local Similarity 85.7%; Pred. No. 7.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 DB 409 LGSVLL 415

RESULT 14  
 ID 053890 PRELIMINARY; PRT; 923 AA.  
 AC 053890;  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
 DE PGRS-FAMILY PROTEIN.  
 GN RV0977 OR MTVO4.05.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxID=1773;  
 RX NCBI\_TaxID=1773;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE-98295987; PubMed-9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 DR EMBL: AL021999; CAI1576.1; -.  
 DR Tuberculol: RV0977; -.  
 DR InterPro: IPR000084; -.  
 DR InterPro: IPR001969; -.  
 DR Pfam: PF00934; PE; 1.  
 DR PRODOM: PD001223; -; 1.  
 DR PROSITE: PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 SQ SEQUENCE 923 AA; 81604 MW; 036E208B6E48C03 CRC64;

Query Match 90.6%; Score 29; DB 2; Length 923;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+03;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 DB 92 VGGVLL 98

RESULT 15  
 ID 09X1K2 PRELIMINARY; PRT; 137 AA.  
 AC 09X1K2;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)

DE HYPOTHETICAL 15.4 KDA PROTEIN.  
GN TM1514.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogales; Thermotoga.  
OX NCBI\_Taxid=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=92287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
genome sequence of Thermotoga maritima."  
RL Nature 399:323-329(1999).  
DR EMBL: AE001799; AAD36581.1; -  
DR TIGR: TM1514; -  
DR InterPro: IPR000139; -  
DR ProDom: PD005242; -, 1.  
KW Hypothetical protein.  
SQ SEQUENCE 137 AA; 15415 MW; 8FB6A95ED9415AEB CRC64;

Query Match 87.5%; Score 28; DB 2; Length 137;  
Best Local Similarity 71.4%; Pred. No. 2.7e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSVLL 7  
:|||||  
Db 64 LGSGILL 70

Search completed: September 6, 2001, 16:49:54  
Job time: 738 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:09 : Search time 72.75 Seconds  
(without alignments)  
3.296 Million cell updates/sec

Title: US-09-603-713-25

Perfect score: 32  
Sequence: 1 VCSGVL 7

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 33435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	96.9	397	1 PURA_THEMA	09x011 thermotoga
2	30	93.8	235	1 AC24_STRCO	P46106 streptomyc
3	29	90.6	116	1 YW02_RHIME	Q52966 rhizobium m
4	29	90.6	274	1 CYSE_BUCAI	P57162 buchnera ap
5	29	90.6	283	1 CRD1_YEAST	007560 saccharomyc
6	29	90.6	431	1 YOD0_MYCTU	006377 mycobacteri
7	29	90.6	510	1 C755_EUSGR	Q96418 eustoma gra
8	29	90.6	510	1 C757_EUSGR	004790 eustoma gra
9	28	87.5	136	1 RRL1_EPIVI	P30059 epifagus vi
10	28	87.5	261	1 CYSE_BUCAP	P32003 buchnera ap
11	28	87.5	293	1 YF09_MYCTU	P71888 mycobacteri
12	28	87.5	354	1 DDL_SYNY3	P73632 synechocyst
13	28	87.5	359	1 OM40_MOUSE	Q99424 mus musculu
14	28	87.5	361	1 OM40_HUMAN	Q92900 homo sapien
15	28	87.5	373	1 DDL_MYCSM	P43525 ascoptillu
16	28	87.5	385	1 GCH2_AZOBH	Q92744 chlamydia p
17	28	87.5	426	1 Y680_CHLPP	P14147 salmonele
18	28	87.5	451	1 PHOO_SAIKY	P08998 escherichia
19	28	87.5	529	1 M8A1_ECOLI	P52019 mus musculu
20	28	87.5	572	1 ERG1_MOUSE	P12522 drosophila
21	28	87.5	584	1 CNA1_DROME	P75291 mycoplasma
22	28	87.5	660	1 SGAT_MYCPN	P37913 mus musculu
23	28	87.5	916	1 DNLI_MOUSE	P18858 homo sapien
24	28	87.5	919	1 DNLI_HUMAN	Q49619 mycobacteri
25	28	87.5	1008	1 MMLA_MYCLE	P38990 saccharomyc
26	28	87.5	1142	1 PAK1_YEAST	Q83889 treponema p
27	27	84.4	105	1 THIO_TREPA	P48230 homo sapien
28	27	84.4	202	1 T4S4_HUMAN	Q66129 aquifex aeo
29	27	84.4	214	1 Y540_AOUAE	Q14027 schizosacch
30	27	84.4	296	1 YEM4_SCHPO	P47698 mycoplasma
31	27	84.4	312	1 LDH_MYCGE	P78007 mycoplasma
32	27	84.4	312	1 LDH_MYCPN	P48325 cyanophora
33	27	84.4	333	1 YC48_CYAPA	

34	27	84.4	354	1 E2B_XYLF	O9p95 xylella fas
35	27	84.4	393	1 YWCF_BACSU	P39604 bacillus su
36	27	84.4	396	1 YWCF_ECOLI	P31442 escherichia
37	27	84.4	403	1 H136_ARATH	O82660 arabidopsis
38	27	84.4	533	1 YN24_YEAST	P48565 saccharomyc
39	27	84.4	573	1 ERG1_RAT	P52020 ratius norv
40	27	84.4	575	1 FLA2_CAMEJ	P22251 campylobact
41	27	84.4	575	1 FLA2_CAMEJ	P22252 campylobact
42	27	84.4	615	1 UAPA_EMENT	O07307 emericella
43	27	84.4	704	1 NEUL_PIG	O02038 sus scrofa
44	27	84.4	2199	1 DPOE_SCHRO	P87154 schizosacch
45	27	84.4	2444	1 NTCL_HUMAN	P46531 homo sapien

## ALIGNMENTS

RESULT ID	PURA_THEMA	STANDARD	PRT	397 AA
AC	09X011:			
DT	30-MAY-2000 (rel. 39, Created)			
DT	30-MAY-2000 (rel. 39, Last sequence update)			
DT	30-MAY-2000 (rel. 39, Last annotation update)			
DE	ADENYLOSUCCLINATE SYNTHETASE (EC 6.3.4.4) (IMP--ASPARTATE LIGASE)			
DE	(ADSS) (AMPSASE).			
GN	PURA OR TMI096.			
OS	Thermotoga maritima.			
OC	Bacteria; Thermotogales; Thermotoga.			
OX	NCBI_TaxID=2336;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MSB / DSM 3109;			
RX	MEDLINE=99287316; PubMed=10360571;			
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;			
RA	*Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.*;			
RT	gene name for lateral gene transfer between Archaea and Bacteria from Nature 399:323-329(1999).			
RT	- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE NUCLEOTIDE BIOSYNTHESIS.			
CC	- CATALYTIC ACTIVITY: GTP + IMP + L-ASPARTATE -> GDP + ORTHOPHOSPHATE + ADENYLOSUCCLINATE.			
CC	- PATHWAY: FIRST COMMITTED STEP IN AMP BIOSYNTHESIS.			
CC	- SIMILARITY: BELONGS TO THE ADENYLOSUCCLINATE SYNTHETASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL: AE001769; AAD36172.1; -			
DR	TIIGR: TMI096; -			
DR	InterPro: IPR001114; -			
DR	Pfam: PF00709; Adenylsucc_synth. 1.			
DR	PROSITE: PS01266; ADENYLOSUCCLIN SYN 1; 1.			
DR	PROSITE: PS00513; ADENYLOSUCCLIN SYN 2; FALSE-NEG.			
KW	Purine biosynthesis; Ligase; GTP-binding			
NP_BIND	11 17			
FT	ACT_SITE 136 136			
FT	ACT_SITE 143 143			
FT	BY SIMILARITY.			
SO	SEQUENCE 397 AA; 44516 MW; FE49BDC5F5B393AD CRC64;			

Query Match 96.9%; Score 31; DB 1; Length 397;  
Best Local Similarity 85.7%; Pred. No. 30;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGSGVLL 7  
:|||||  
Db 67 IGSGVLL 73

RESULT 2  
AC24\_STRCO STANDARD; PRT; 255 AA.  
AC P46106;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DE 01-NOV-1997 (Rel. 35, Last annotation update)  
DE PROBABLE ACTINORHODIN OPERON ACTIVATORY PROTEIN.  
GN ACTII-4.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91347376; PubMed=1878971;  
RA Fernandez-Moreno M.A., Caballero J.L., Hopwood D.A., Malpartida F.;  
RT "The act cluster contains regulatory and antibiotic export genes,  
RT direct targets for translational control by the bla<sub>TRNA</sub> gene of  
RT Streptomyces";  
RL Cell 66:769-780(1991).  
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF ACTINORHODIN BIOSYNTHETIC  
CC GENES.  
CC -1- SIMILARITY: BELONGS TO THE AFSR/DNRI/REDD FAMILY OF REGULATORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M64683; AAA26692.1; -  
CC Antibiotic biosynthesis; Transcription regulation; Activator.  
KW AMINO ACID SEQUENCE 255 AA; 28762 MW; 59E1E7A9C37F3AD CRC64;  
SQ SEQUENCE

Query Match 93.8%; Score 30; DB 1; Length 255;  
Best Local Similarity 85.7%; Pred. No. 31;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGSGVLL 7  
:|||||  
Db 41 VGSGVLL 47

RESULT 3  
YMO2\_RHME STANDARD; PRT; 116 AA.  
AC Q52966;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HYPOTHETICAL 12.4 KDa PROTEIN IN MUCR 5' REGION (ORF2).  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RCR2011 / SUA7;  
RX MEDLINE=95276304; PubMed=7756693;  
RA Keller M., Koxland A., Weng W.M., Schmidt M., Quandt J., Niehaus K.,  
RA Jording D., Arnold W., Puhler A.;  
RT "Molecular analysis of the Rhizobium meliloti mucr gene regulating the

RT biosynthesis of the exopolysaccharides succinoglycan and  
RT galactoglucan";  
RT Mol. Plant Microbe Interact. 8:267-277(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021.  
RA Galliard F., Capela D., Hubler-Barloy F., Gattus M., Batut J.,  
RA Bolstad P., Guzy J., Kahn D., Thebaud P., Goffeau A.,  
RA Purnelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,  
RA Vandenbol M., Puhler A., Becker A., Weidner S.;  
RL Submitted (MAR-2000) to the SWISS-PROT data bank.  
CC -1- SIMILARITY: BELONGS TO THE UPF0091 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U37353; AAA74238.1; -  
CC Hypothetical protein.  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 116 AA; 12373 MW; 16F4C117B81C4D10 CRC64;  
SQ SEQUENCE

Query Match 90.6%; Score 29; DB 1; Length 116;  
Best Local Similarity 85.7%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGSGVLL 7  
:|||||  
Db 47 VGSGVLL 53

RESULT 4  
CYSE\_BUCAI STANDARD; PRT; 274 AA.  
AC P57162;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE SERINE ACETYLTANSFERASE (EC 2.3.1.30) (SAT).  
GN CYSE OR BU054.  
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
OS symbiotic bacterium).  
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.  
OX NCBI\_TaxID=118099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TOKYO 1998;  
RX MEDLINE=20445173; PubMed=10993077;  
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
RT "Genome sequence of the endocellular bacterial symbiont of aphids  
RT Buchnera sp. APS.";  
RL Nature 407:81-86(2000).  
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + L-SERINE -> COA + O-ACETYL-L-  
CC SERINE.  
CC -1- PATHWAY: CYSTEINE BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NDL FAMILY OF  
CC ACETYLTANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AP001118; BAB12777.1; -  
CC PROSITE: PS00101; HEXAPEP\_TRANSFERRASES; 1.  
DR PROSITE

KW Transferrase; Acyltransferase; Cysteine biosynthesis; Repeat.  
SQ SEQUENCE 274 AA; 30165 MW; 553D252F1048B6B1 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 274;  
Best Local Similarity 71.4%; Pred. NO. 54;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGVLL 7  
:|||||:  
DB 150 ICGSVML 156

## RESULT 5

CRD1\_YEAST STANDARD; PRT; 283 AA.  
ID CRD1\_YEAST  
AC 007560; P82260; (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DR 30-MAY-2000 (Rel. 39, Last annotation update)  
DE CARDIOLIPIN SYNTHETASE (EC 2.7.8.-) (CARDIOLIPIN SYNTHASE) (CLS).  
GN CRD1 OR CLS1 OR YDL142C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Barton L., Legros Y., Biteau N., Monnet A., Granotier C.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 106-135 FROM N.A.  
RA Sarma P.V.G.K., Sarma P.U.;  
RL Submitted (DEC-1999) to the SWISS-PROT data bank.  
RN [3]  
RP IDENTIFICATION.  
RA MEDLINE=96122662; PubMed=9462830;  
RA Tuller G., Hrstnik C., Acheltner G., Schieffeler U., Klein F.,  
RA Daum G.;  
RL "YDL142C encodes cardiolipin synthase (CLS1) and is non-essential for  
aerobic growth of Saccharomyces cerevisiae.";  
RL FEBS Lett. 421:15-18(1998).  
RN [4]  
RP IDENTIFICATION.  
RA MEDLINE=98279006; PubMed=9614098;  
RA Chang S.C., Heacock P.N., Mileykovskaya E., Voelker E.R., Dowhan W.;  
RL "Isolation and characterization of the gene (CLS1) encoding  
cardiolipin synthase in Saccharomyces cerevisiae.";  
RL J. Biol. Chem. 273:14933-14941(1998).  
RN [5]  
RP CHARACTERIZATION.  
RA MEDLINE=98036101; PubMed=9370334;  
RA Schlame M., Greenberg M.L.;  
RL "Cardiolipin synthase from yeast.";  
RL Biochim. Biophys. Acta 1348:201-206(1997).  
CC -1- FUNCTION: CATALYZES THE REVERSIBLE PHOSPHATIDYL GROUP TRANSFER  
FROM ONE PHOSPHATIDYLGLYCEROL MOLECULE TO ANOTHER TO FORM  
CARDIOLIPIN (CL) (BIPHOSPHATIDYLGLYCEROL) AND GLYCEROL.  
CC -1- CATALYTIC ACTIVITY: 2 PHOSPHATIDYLGLYCEROL =  
BIPHOSPHATIDYLGLYCEROL + GLYCEROL.  
CC -1- COPACITOR: REDUCES A DIVALENT ION FOR ACTIVITY.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
INNER MEMBRANE.  
CC -1- SIMILARITY: BELONGS TO THE CDP-ALCOHOL PHOSPHATIDYLTRANSFERASE  
CLASS-1 FAMILY.  
CC -1- CAUTION: REF.2 SEQUENCE WAS ORIGINALLY THOUGHT TO ORIGINATE FROM  
MYCOBACTERIUM PHLEI.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: 274190; CAA98715.1; -  
DR SGD: S0002301; CRD1.  
DR InterPro: IPR000462; -  
DR Pfam: PF01066; CDP-OR\_P\_transf. 1.  
DR ProSite: PS00379; CDP\_ALCOHOL\_P\_TRANSF. 1.  
KW Transferrase; Phospholipid biosynthesis; Transmembrane; Mitochondrion.  
FT TRANSMEM 83 103  
FT TRANSMEM 155 175  
FT TRANSMEM 209 229  
FT TRANSMEM 283 AA; 32019 MW; F19F4FE8D9BE049 CRC64;  
SQ SEQUENCE

Query Match 90.6%; Score 29; DB 1; Length 283;  
Best Local Similarity 85.7%; Pred. NO. 56;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGVLL 7  
:|||||:  
DB 221 ICGSVLL 227

## RESULT 6

YOD0\_MYCTU STANDARD; PRT; 431 AA.  
ID YOD0\_MYCTU  
AC 006377;  
DT 01-NOV-1997 (Rel. 35, Created)  
DR 01-NOV-1997 (Rel. 35, Last sequence update)  
DE 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HYPOTHETICAL 43.5 KDA PROTEIN RV3630.  
GN RV3630 OR MYCY15C10.22C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=H37RV;  
RC MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,  
RA Horsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skellton S., Squares S., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RL "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
RL Nature 393:537-544(1998).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV1510.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: 295436; CAB08850.1; -  
DR Tuberculist: RV3630; -  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 33 53  
FT TRANSMEM 63 83  
FT TRANSMEM 111 131  
FT TRANSMEM 143 163  
FT TRANSMEM 175 195  
FT TRANSMEM 197 217  
FT TRANSMEM 241 261  
FT TRANSMEM

FT TRANSMEM 273 293 POTENTIAL.  
 FT TRANSMEM 318 338 POTENTIAL.  
 FT TRANSMEM 358 378 POTENTIAL.  
 FT TRANSMEM 381 401 POTENTIAL.  
 FT TRANSMEM 407 427 POTENTIAL.  
 SO SEQUENCE 431 AA; 43485 MW; 34CFB2876700114 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 431;  
 Best Local Similarity 85.7%; Pred. No. 85;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 11111111  
 Db 391 VSGSVLL 397

RESULT 7  
 C755\_EUSGR STANDARD; PRT; 510 AA.  
 AC 096418;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE FLAVONOID 3',5'-HYDROXYLASE (EC 1.14.-.-) (F3'5'H) (CYTOCHROME P450 75A5).  
 GN CYP75A5.  
 OS Eustoma grandiflorum (Bluebells) (Lisianthus russellianus).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 CC Gentianales; Gentianaceae; Eustoma.  
 RN NCBI\_TaxID=52518;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Petal;  
 RA Nielsen K.M., Podivinsky E.;  
 RT "CDNA Cloning and Expression of a flavonoid 3',5'-hydroxylase from  
 petals of Lisianthus (Eustoma grandiflorum Grise).";  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CATALYSES THE 3',5'-HYDROXYLATION OF NARINGENIN AND  
 ERIDICTYOL TO FORM 5,7,3',4',5'-PENTAHYDROXYFLAVANONE AND 3',5'-  
 HYDROXYLATION OF DIHYDROKEMPEROL AND DIHYDROQUERCETIN TO FORM  
 DIHYDROMYRICETIN (BY SIMILARITY).  
 CC -1- PATHWAY: ANTHOCYANIN BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U72654; AAB17562.1; -;  
 DR InterPro; IPR001128; -;  
 DR Pfam; PF00067; P450; 1.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KM Oxidoreductase; Monooxygenase; Heme.  
 FT BINDING 447 447 HEME (BY SIMILARITY).  
 GN BINDING 447 447 HEME (BY SIMILARITY).  
 SO SEQUENCE 510 AA; 56894 MW; 8B6AB02DD9A5BAB9 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 510;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 11111111  
 Db 3 VSGSVLL 9

RESULT 8

C757\_EUSGR STANDARD; PRT; 510 AA.  
 ID C757\_EUSGR  
 AC 004790;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE FLAVONOID 3',5'-HYDROXYLASE (EC 1.14.-.-) (F3'5'H) (CYTOCHROME P450 75A7).  
 GN CYP75A7.  
 OS Eustoma grandiflorum (Bluebells) (Lisianthus russellianus).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 CC Gentianales; Gentianaceae; Eustoma.  
 RN NCBI\_TaxID=52518;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ROYAL LIGHT PURPLE; TISSUE=Petal;  
 RA Ohbayashi M.;  
 RL Submitted (MAR-1993) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CATALYSES THE 3',5'-HYDROXYLATION OF NARINGENIN AND  
 ERIDICTYOL TO FORM 5,7,3',4',5'-PENTAHYDROXYFLAVANONE AND 3',5'-  
 HYDROXYLATION OF DIHYDROKEMPEROL AND DIHYDROQUERCETIN TO FORM  
 DIHYDROMYRICETIN (BY SIMILARITY).  
 CC -1- PATHWAY: ANTHOCYANIN BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D14589; BAA03439.1; -;  
 DR InterPro; IPR001128; -;  
 DR Pfam; PF00067; P450; 1.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KM Oxidoreductase; Monooxygenase; Heme.  
 FT BINDING 447 447 HEME (BY SIMILARITY).  
 GN BINDING 447 447 HEME (BY SIMILARITY).  
 SO SEQUENCE 510 AA; 56732 MW; B0A84BF2DB2771D CRC64;

Query Match 90.6%; Score 29; DB 1; Length 510;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 11111111  
 Db 3 VSGSVLL 9

RESULT 9  
 R11\_EPIVI STANDARD; PRT; 136 AA.  
 ID R11\_EPIVI  
 AC P30059;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE CHLOROPLAST 30S RIBOSOMAL PROTEIN S11.  
 GN RPS11.  
 OS Epifagus virginiana (Beechdrops).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 CC Lamiales; Scrophulariaceae; Epifagus.  
 RN NCBI\_TaxID=4177;  
 [1]  
 RP SEQUENCE FROM N.A.  
 KM MEDLINE=93066301; PubMed=1332054;  
 RX Wolfe K.H., Morden C.W., Palmer J.D.;  
 RT "Function and evolution of a minimal plastid genome from a  
 nonphotosynthetic parasitic plant.";

```

RL Proc. Natl. Acad. Sci. U.S.A. 89:10648-10652(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-93021155; PubMed-1404416;
RA Wolfe K.H., Morden C.W., Ems S.C., Palmer J.D.:
RT "Rapid evolution of the plastid translational apparatus in a
nonphotosynthetic plant: loss or accelerated sequence evolution of
tRNA and ribosomal protein genes.";
RL J. Mol. Evol. 35:304-317(1992).
CC -1- SIMILARITY: BELONGS TO THE SLP FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL: M81884; AAA65859.1; -.
CC InterPro: IPR001971; -.
CC Pfam: PF00411; Ribosomal_S11; 1.
CC PROSITE: PS00054; RIBOSOMAL_S11; 1.
CC Ribosomal protein: Chloroplast.
CC RIBOSOMAL PROTEIN: Chloroplast.
CC SEQUENCE 136 AA; 14650 MW; 57D02B67740F6E9D CRC64;

Query Match 87.5%; Score 28; DB 1; Length 136;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GSGVLL 7
DB 109 GSGVLL 114

RESULT 10
CYSE_BUCAP STANDARD; PRT; 261 AA.
AC P32003;
DT 01-JUL-1993 (Rel. 26; Created)
DT 01-JUL-1993 (Rel. 26; Last sequence update)
DT 30-MAY-2000 (Rel. 39; Last annotation update)
DE SERINE ACETYLTRANSFERASE (EC 2.3.1.30) (SAT).
GN CYSE.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93012960; PubMed-1398077;
RA Lal C.Y., Baumann P.:
RT "Sequence analysis of a DNA fragment from Buchnera aphidicola (an
endosymbiont of aphids) containing genes homologous to dnaG, rpoD,
rt cysE, and secB.";
RL Gene 119:113-118(1992).
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + L-SERINE -> COA + O-ACETYL-L-
SERINE.
CC -1- PATHWAY: CYSTEINE BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NODL FAMILY OF
ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL: M90644; AAA73232.1; -.

```

```

DR PIR: JC1293; JC1293.
DR InterPro: IPR001451; -.
DR Pfam: PF00132; hexaped; 2.
DR PROSITE: PS00101; HEXAPED_TRANSFERASES; 1.
KW TRANSFERASE; Acyltransferase; Cysteine biosynthesis; Repeat.
SQ SEQUENCE 261 AA; 28914 MW; DC02685464A25944 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 261;
Best Local Similarity 57.1%; Pred. No. 84;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGVLL 7
DB 150 IGSGLML 156

RESULT 11
YF09_MYCTU STANDARD; PRT; 293 AA.
ID YF09_MYCTU
AC P71788;
DT 01-NOV-1997 (Rel. 35; Created)
DT 01-NOV-1997 (Rel. 35; Last sequence update)
DT 30-MAY-2000 (Rel. 39; Last annotation update)
DE HYPOTHETICAL 33.0 KDA PROTEIN RV1509.
GN RV1509 OR MTCY277.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RC MEDLINE-98295987; PubMed-9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squires S., Sulton J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL: 279701; CAB02023.1; -.
CC DR TubercuList; RV1509; -.
CC KW Hypothetical protein.
SQ SEQUENCE 293 AA; 33055 MW; 7393B5725FE71EE7 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 293;
Best Local Similarity 85.7%; Pred. No. 94;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 VSGVLL 7
DB 67 VGGGVLL 73

RESULT 12
DDL_SYNY3 STANDARD; PRT; 354 AA.
ID DDL_SYNY3
AC P73632;

```

```

DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      D-ALANINE--D-ALANINE LIGASE (EC 6.3.2.4) (D-ALANYLALANINE SYNTHETASE)
DE      (D-ALA-D-ALA LIGASE).
GN      DDL OR DDLA OR SLR1874.
OS      Synchocystis sp. (strain PCC 6803).
OC      Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX      NCBI_TaxId=1148;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97061201; PubMed=8905231.
RA      Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA      Miyajima N., Hiroseawa M., Sugitara M., Sasamoto S., Kimura T.,
RA      Hosouchi T., Matsuno A., Muraiki A., Nakazaki N., Naruo K.,
RA      Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A.,
RA      Yamada M., Yasuda M., Tabata S.;
RT      "Sequence analysis of the genome of the unicellular cyanobacterium
RT      Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT      entire genome and assignment of potential protein-coding regions.";
RL      DNA Res. 3:109-136(1996).
CC      -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: ATP + D-ALANINE + D-ALANINE = ADP +
CC      ORTHOPHOSPHATE + D-ALANYL-D-ALANINE.
CC      -1- PATHWAY: ALONG WITH ALANINE RACEMASE, IT MAKES UP THE D-ALANINE
CC      BRANCH OF THE PEPTIDOGLICAN BIOSYNTHETIC ROUTE (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE D-ALANINE--D-ALANINE LIGASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isp.slb.ch/announce/
CC      or send an email to license@isp.slb.ch).
CC      -----
DR      EMBL: D80908; BAA17677.1; -
DR      HSSP: P07862; 2DLN.
DR      InterPro: IPR000291; -
DR      Pfam: PF01820; Data_Dala_Ligas: 1.
DR      PROSITE: PS00843; DATA_DALA_LIGASE_1: 1.
DR      PROSITE: PS00844; DATA_DALA_LIGASE_2: 1.
KW      Ligase; Cell wall; Peptidoglycan synthesis.
SQ      SEQUENCE 354 AA; 38774 MW; 871B82362A47315A CRC64;

Query Match          87.5%; Score 28; DB 1; Length 354;
Best Local Similarity 100.0%; Pred. No. 1;le=02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Oy      1 VSGVL 6
Db      115 VSGVL 120

RESULT 13
OM40_MOUSE
ID      OM40_MOUSE          STANDARD;          PRT;          359 AA.
AC      O9OYAZ2_Q9Z2N1;
DT      01-OCT-2000 (Rel. 40, Created)
DT      01-OCT-2000 (Rel. 40, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      (PROBABLY MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM40 HOMOLOG
DE      (TOM40LOCUS OF OUTER MEMBRANE 40 KDA SUBUNIT TOMO40)).
GN      TOM40. OR TOM40 OR MOM35.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxId=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Lee Rivera I., Schlieff E., Shore G.C.;

```

```

RT *Cloning and characterization of a mitochondrial protein with high
RT homology to Tom40."
RL J. Bioenerg. Biomembr. 0:0-0(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Yoshitura K., Murray J.C.;
RT "A transcriptional map in the region of 19q13 derived using direct
RT sequencing and exon trapping".
CC Submitted (JAN-1998) to the EMBL/GenBank/DDBB databases.
CC -1- FUNCTION: ESSENTIAL FOR THE IMPORT OF PROTEIN PRECURSORS INTO THE
CC MITOCHONDRIA (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC OUTER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TOM40 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isdb.ch/announce/
CC or send an email to license@isdb.sib.ch).
-----
CC -----
DR EMBL; AF109918; AAF21906.1; -
DR EMBL; AF043249; AAC82341.1; -
DR MDD; MG1:1688259; Tomm40.
KW Transport; Protein transport; Outer membrane; Mitochondrion;
KM Transmembrane.
FT CONFLICT 1 76
FT GNLYASPPARPPTSTVLGLPDPSPSSPLCRSAA
FT LGTSSGSRGSEPTPGAAASAEDESCCLP -> NM
FT KSGDMVHMHWPMFGTDNRGGCGRRCGGLGRWELRMPA (IN
FT REF. 1).
FT MISSING (IN REF. 2).
FT FT FT MISSING IN REF. 2).
FT FT FT MISSING IN REF. 2).
SQ SEQUENCE 359 AA; 38001 MW; A203DIFSC0678003 CRC64;

QY 1 VGSGVLL 7
   |||||:
Db 210 VGSGLIV 216

Query Match      87.5%; Score 28; DB 1; Length 359;
Best Local Similarity 71.4%; Pred. NO. 1.2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
ID OM40_HUMAN STANDARD; PRT; 361 AA.
AC O96008;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROBABLE MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM40 HOMOLOG
DE (TRANSLOCASE OF OUTER MEMBRANE 40 KDA SUBUNIT HOMOLOG).
OS Homo sapiens (Human).
GN TOMM40 OR TOM40 OR PEREC1.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mamalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=9449047; Pubmed=10520737;
RA Freitas E.M., Zhang W.J., Lalonde J.P., Tay G.K., Gaudieri S.,
RA Ashworth L.K., Van Bockmeier F.M., Dawkins R.L.;
RT "Sequencing of 42kb of the APO E-C2 gene cluster reveals a new gene:
RT PEREC1."
RL DNA Seq. 9:89-101(1998).
RN (2)
RP SEQUENCE FROM N.A.
RA Yoshitura K., Murray J.C.;
RT "A transcriptional map in the region of 19q13 derived using direct
RT sequencing and exon trapping".
RL Submitted (JAN-1998) to the EMBL/GenBank/DDBB databases.
```

CC -1- FUNCTION: ESSENTIAL FOR THE IMPORT OF PROTEIN PRECURSORS INTO THE  
 CC MITOCHONDRION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC OUTER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE TOM40 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF050154; AAC02504.1; -  
 CC EMBL; AF043250; AAC82342.1; -  
 CC EMBL; AF043253; AAC82343.1; -  
 CC EMBL; AF043251; AAC82343.1; JOINED.  
 CC EMBL; AF043252; AAC82343.1; JOINED.  
 CC Transport; Protein transport; Outer membrane; Mitochondrion;  
 CC Transmembrane.  
 CC  
 CC SEQUENCE 361 AA; 37893 MW; CFE5E01F8003D32 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 361;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVL 7  
 |||||:  
 Db 212 VSGSVL 218

RESULT 15  
 DDL\_MYCSM STANDARD; PRT; 373 AA.  
 ID DDL\_MYCSM  
 AC Q9ZGN0;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE D-ALANINE--D-ALANINE LIGASE (EC 6.3.2.4) (D-ALANYLALANINE SYNTHETASE)  
 DE (D-ALA-D-ALA LIGASE).  
 DDL.  
 GN  
 OS Mycobacterium smegmatis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1772;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Belanger A.E., Porter J.C., Hatfull G.F.;  
 RT "A temperature-sensitive mutant of Mycobacterium smegmatis which is  
 RT defective in cell wall biosynthesis."  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + D-ALANINE + D-ALANINE -> ADP +  
 CC ORTHOPHOSPHATE + D-ALANYL-D-ALANINE.  
 CC -1- PATHWAY: ALONG WITH ALANINE RACEMASE, IT MAKES UP THE D-ALANINE  
 CC BRANCH OF THE PEPTIDOGLYCAN BIOSYNTHETIC ROUTE.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE D-ALANINE--D-ALANINE LIGASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF077728; AAC99396.1; -  
 CC HSSP; P07862; 2DLN.  
 CC InterPro: IPR000291; -  
 CC Pfam: PF01820; Data\_Ligas; 1.

DR PROSITE: PS00843; DATA\_DATA\_LIGASE\_1; 1.  
 DR PROSITE: PS00844; DATA\_DATA\_LIGASE\_2; 1.  
 KW Ligase; Cell wall; Peptidoglycan synthesis.  
 SO SEQUENCE 373 AA; 39231 MW; FE7151D858C39D43 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 373;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSVL 6  
 |||||  
 Db 139 VSGSVL 144

Search completed: September 6, 2001, 16:51:10  
 Job time: 814 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:52 : Search time 134.15 Seconds  
(without alignments)  
3.975 Million cell updates/sec

Title: US-09-603-713-25

Perfect score: 32

Sequence: 1 VSGSVLL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR\_68:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	96.9	397	2 B72294	adenylosuccinate s
2	30	93.8	227	2 D84058	transcription prote
3	30	93.8	255	2 D40046	di-tripeptide ABC
4	30	93.8	510	2 H82624	hypothetical prote
5	29	90.6	80	2 T29969	conserved hypothet
6	29	90.6	82	2 D72280	lipopolysaccharide
7	29	90.6	162	2 C82738	serine O-acetyltra
8	29	90.6	274	2 A84936	probable membrane
9	29	90.6	283	2 S67889	hypothetical prote
10	29	90.6	375	2 T04036	hypothetical prote
11	29	90.6	421	2 H96550	hypothetical prote
12	29	90.6	431	2 G70561	hypothetical prote
13	29	90.6	442	2 H70960	hypothetical prote
14	29	90.6	547	2 T35267	probable ABC trans
15	29	90.6	923	2 E70820	hypothetical glyci
16	28	87.5	136	2 S78390	ribosomal protein
17	28	87.5	137	2 C72245	hypothetical prote
18	28	87.5	144	2 S66938	probable membrane
19	28	87.5	149	2 H75464	probable cation tr
20	28	87.5	224	2 T35918	hypothetical prote
21	28	87.5	261	1 JCI293	serine O-acetyltra
22	28	87.5	293	2 H70713	hypothetical prote
23	28	87.5	306	2 T29337	hypothetical prote
24	28	87.5	319	2 G83493	hypothetical prote
25	28	87.5	326	2 T36421	hypothetical prote
26	28	87.5	339	2 D75529	D-alanine--D-alani
27	28	87.5	354	2 S77119	D-alanine--D-alani
28	28	87.5	382	2 D82295	npcA protein VC067
29	28	87.5	385	1 I39498	GTP cyclohydrolase

30	28	87.5	395	2 T44985	methylnalonyl-CoA
31	28	87.5	416	2 G81258	serine transporter
32	28	87.5	426	2 E86575	phosphate permease
33	28	87.5	426	2 D72049	phosphate permease
34	28	87.5	438	2 A71323	hypothetical prote
35	28	87.5	444	2 T24204	hypothetical prote
36	28	87.5	451	2 E83418	probable amino aci
37	28	87.5	462	2 F83510	flagellar hook pro
38	28	87.5	463	2 B82001	probable amino-aci
39	28	87.5	463	2 B81228	sodium/alanine sym
40	28	87.5	487	1 VZEBPT	sensor kinase phoC
41	28	87.5	493	2 T29833	hypothetical prote
42	28	87.5	516	2 T29834	hypothetical prote
43	28	87.5	529	1 MZEC8	58k mobilization p
44	28	87.5	572	2 S52113	squalene epoxidase
45	28	87.5	579	2 JCI7170	fimbria-like 71 k

#### ALIGNMENTS

##### RESULT 1

B72294 adenylosuccinate synthetase - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: B72294

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.

Nature 399, 323-329, 1999

A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316

A:Accession: B72294

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-397 <ARN>

A:Cross-references: GB:AE001769; GB:AE000512; NID:94981639; PIDN:AAD36172.1; PID:9498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1096

C:Superfamily: adenylosuccinate synthase

Query Match 96.9%; Score 31; DB 2; Length 397;

Best Local Similarity 85.7%; Pred. No. 48;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSVLL 7

DB 67 IGSVLL 73

##### RESULT 2

D84058 hypothetical protein BH3268 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000

C:Accession: D84058

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20263314

A:Accession: D84058

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-227 <STO>

A:Cross-references: GB:AP001518; GB:BA000004; NID:910175792; PIDN:BA06987.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3268

C:Superfamily: yohk protein

Query Match 93.8%; Score 30; DB 2; Length 227;  
 Best Local Similarity 85.7%; Pred. No. 45;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 |||||  
 Db 102 VSGSVLM 108

RESULT 3  
 D40046  
 transcription activator acti-4 - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 22-Oct-1999  
 C:Accession: D40046  
 R.Fernandez-Moreno, M.A.; Caballero, J.L.; Hopwood, D.A.; Malpartida, F.  
 Cell 66, 769-780, 1991  
 A:Title: The act cluster contains regulatory and antibiotic export genes, direct targets  
 A:Reference number: M40046; MUID:91347376  
 A:Accession: D40046  
 A:Molecule type: DNA  
 A:Residues: 1-255 <FER>  
 A:Cross-references: GB:M64683; NID:g153143; PIDN:AAA26692.1; PID:g153147

Query Match 93.8%; Score 30; DB 2; Length 255;  
 Best Local Similarity 85.7%; Pred. No. 50;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 |||||  
 Db 41 VSGSVLL 47

RESULT 4  
 H82624  
 di-tripeptide ABC transporter membrane protein Xf1891 [imported] - Xylella fastidiosa (S  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
 C:Accession: H82624  
 R.anonymus, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: H82624  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-510 <SIM>  
 A:Cross-references: GB:AE004009; GB:AE003849; NID:g9106980; PIDN:AAF84697.1; GSPDB:GN001  
 A:Experimental source: strain 9a5C  
 R.Stimpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigi  
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tshakro, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: Xf1891  
 C:Superfamily: peptide transporter protein

Query Match 93.8%; Score 30; DB 2; Length 510;  
 Best-Local Similarity 71.4%; Pred. No. 1e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :|||:  
 Db 489 IGSGLL 495

RESULT 5  
 T29969  
 hypothetical protein ZK682.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T29969  
 R.Du, Z.; Le, T.  
 submitted to the EMBL Data Library, November 1995  
 A:Description: The sequence of C. elegans cosmid ZK682.  
 A:Reference number: Z20714  
 A:Accession: T29969  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-80 <DUZ>  
 A:Cross-references: EMBL:U41110; PIDN:AA82416.1; CESP:ZK682.3  
 A:Gene: CESP:ZK682.3  
 A:Introns: 64/2

Query Match 90.6%; Score 29; DB 2; Length 80;  
 Best Local Similarity 85.7%; Pred. No. 25;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 |||||  
 Db 22 VSGSVLL 28

RESULT 6  
 D72280  
 conserved hypothetical protein - Thermotoga maritima (strain MS88)  
 C:Species: Thermotoga maritima  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C:Accession: D72280  
 R.Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
 Garrell, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
 C.M.  
 Nature 399, 323-329, 1999  
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
 A:Reference number: A72200; MUID:99287316  
 A:Accession: D72280  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-82 <RRN>  
 A:Cross-references: GB:AE001778; GB:AE000512; NID:g4981757; PIDN:AAD36281.1; PID:g498  
 A:Experimental source: strain MS88  
 C:Genetics:  
 A:Gene: TM1206

Query Match 90.6%; Score 29; DB 2; Length 82;  
 Best Local Similarity 85.7%; Pred. No. 26;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 |||||  
 Db 9 VGAGVLL 15

RESULT 7  
 C82738  
 lipopolysaccharide synthesis enzyme Xf0980 [imported] - Xylella fastidiosa (strain 9a  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
 C:Accession: C82738

Anonymous. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: C87738  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-162 <SIM>  
 A:Cross-references: GB:AE003936; GB:AE003849; NID:g9105908; PIDN:AAF83790.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Stimpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H. as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B. A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y., F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiter, D.A. Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF0980  
 C:Superfamily: lipopolysaccharide core biosynthesis protein kdtB

Query Match 90.6%; Score 29; DB 2; Length 162;  
 Best Local Similarity 85.7%; Pred. No. 52;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :|||||  
 DB 85 VSGAVLL 91

RESULT 8  
 A84936  
 serine O-acetyltransferase (EC 2.3.1.30) [imported] - Buchnera sp. (strain APS)  
 C:Species: Buchnera sp.  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
 C:Accession: A84936  
 R:Shigenobu, S.; Matanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.  
 Nature 407, 81-86, 2000  
 A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A  
 A:Reference number: A84930; MUID:20445173  
 A:Accession: A84936  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-274 <STO>  
 A:Cross-references: GB:AP000398; GSPDB:GN00144  
 A:Experimental source: strain APS  
 C:Genetics:  
 A:Gene: cysE; BU054  
 C:Superfamily: serine acetyltransferase; serine acetyltransferase homology  
 C:Keywords: acyltransferase; coenzyme A

Query Match 90.6%; Score 29; DB 2; Length 274;  
 Best Local Similarity 71.4%; Pred. No. 90;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :|||||  
 DB 150 IGSVWL 156

RESULT 9  
 S67689  
 Probable membrane protein YDL142c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein D2000  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 06-Feb-1998  
 C:Accession: S67689  
 R:Baron, L.; Legros, Y.; Biteau, N.; Monnet, A.; Granotier, C.  
 submitted to the Protein Sequence Database, July 1996  
 A:Reference number: S67688  
 A:Accession: S67689  
 A:Molecule type: DNA  
 A:Residues: 1-283 <BAR>  
 A:Cross-references: EMBL:Z74190; NID:g1431220; PID:q253057; PID:g1431221; MIPS:YDL142  
 A:Experimental source: strain S288C  
 C:Genetics:  
 A:Gene: SGD:PGS1  
 A:Cross-references: SGD:S0002301; MIPS:YDL142c  
 A:Map position: 4L  
 C:Keywords: transmembrane protein  
 F:73-69/Domain: transmembrane #status predicted <TM1>  
 F:159-175/Domain: transmembrane #status predicted <TM2>

Query Match 90.6%; Score 29; DB 2; Length 283;  
 Best Local Similarity 85.7%; Pred. No. 93;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :|||||  
 DB 221 IGSVLL 227

RESULT 10  
 T04036  
 hypothetical protein F24G24.10 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 01-Dec-2000  
 C:Accession: T04036; T04071  
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K. submitted to the Protein Sequence Database, March 1999  
 A:Reference number: Z15184  
 A:Accession: T04036  
 A:Molecule type: DNA  
 A:Residues: 1-375 <BEV>  
 A:Cross-references: EMBL:AL049488  
 A:Experimental source: cultivar Columbia; BAC clone F24G24  
 A:Accession: T04071  
 A:Molecule type: DNA  
 A:Residues: 1-173, 'VSAFL' <BEZ>  
 A:Cross-references: EMBL:AL049487  
 A:Experimental source: cultivar Columbia; BAC clone F28M11  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 64/3; 95/3; 173/3; 227/1; 266/3  
 A:Note: F24G24.10; F28M11.130

Query Match 90.6%; Score 29; DB 2; Length 375;  
 Best Local Similarity 71.4%; Pred. No. 1,2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :|||||  
 DB 238 VSGSVLL 244

RESULT 11  
 H96550  
 hypothetical protein F11M15.16 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
 C:Accession: H96550  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizart, L.

Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziani, R.; Rooney, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A66141; MUID:21016719  
 A:Accession: H95550  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-421 <STO>  
 A:Cross-references: GB:AE005173; NID:94836940; PIDN:AD030642.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: P1LM15.16  
 A:Map position: 1  
 C:Superfamily: probable membrane protein YDL033c

Query Match 90.6%; Score 29; DB 2; Length 421;  
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :||||:  
 Db 358 IGSVLL 364

## RESULT 12

hypothetical protein RV3630 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: G70561  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: G70561  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-431 <COL>  
 A:Cross-references: GB:295436; GB:AL123456; NID:93261770; PIDN:CAB08850.1; PID:92105064  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV3630

Query Match 90.6%; Score 29; DB 2; Length 431;  
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :||||:  
 Db 391 VSGSVLL 397

## RESULT 13

hypothetical protein RV0218 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: H70960  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987

A:Accession: H70960  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-442 <COL>  
 A:Cross-references: GB:292669; GB:AL123456; NID:93242271; PIDN:CAB07013.1; PID:930495  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV0218

Query Match 90.6%; Score 29; DB 2; Length 442;  
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :||||:  
 Db 253 VSGSVLL 259

## RESULT 14

probable ABC transporter - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
 C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 17-Mar-2000  
 C:Accession: T35267  
 R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, April 1999  
 A:Reference number: Z21573  
 A:Accession: T35267  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-547 <COL>  
 A:Cross-references: EMBL:AL049587; PIDN:CAB40692.1; GSPDB:GN00070; SCOEDB:SC5F2A.25  
 C:Genetics:  
 A:Gene: SCOEDB:SC5F2A.25  
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match 90.6%; Score 29; DB 2; Length 547;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 :||||:  
 Db 409 IGSVLL 415

## RESULT 15

hypothetical glycine-rich protein RV0977 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C:Accession: E70820  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: E70820  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-923 <COL>  
 A:Cross-references: GB:AL021999; GB:AL123456; NID:93261538; PIDN:CA17576.1; PID:9291  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV0977  
 C:Superfamily: elastin

Query Match 90.6%; Score 29; DB 2; Length 923;

Best Local Similarity 85.7%, Pred. No. 3.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 VSGGVLL 7  
11:1111  
Db 92 VSGGVLL 98

Search completed: September 6, 2001, 16:45:53  
Job time: 497 sec

---





US-08-606-505B-66

Query Match 90.6%; Score 29; DB 3; Length 510;  
Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGSVLL 7  
|:||||  
DB 3 VGNVLL 9

## RESULT 2

US-09-616-990-66  
; Sequence 66, Application US/09616990  
; Patent No. 6232109  
; GENERAL INFORMATION:  
; APPLICANT: KIKUCHI, Yasuhiro  
; KIKUCHI, Shigeto  
; SHIMADA, Yukihisa  
; OHBAYASHI, Masaya  
; SHIMADA, Ritsuko  
; OKINAKA, Yasushi  
; TITLE OF INVENTION: NOVEL PLANT GENES  
; NUMBER OF SEQUENCES: 67  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FITZPATRICK, CELLA, HARPER & SCINTO  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10112-3801  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: diskette - 3.50 inch, 720 Kb storage.  
; OPERATING SYSTEM: IBM PS/VS  
; SOFTWARE: PATENT AID Ver1.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/616,990  
; FILING DATE: 14-Jul-2000  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: JP44963/92  
; FILING DATE: 02-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perry, Lawrence S.  
; REGISTRATION NUMBER: 31865  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-218-2100  
; TELEFAX: 212-218-2200  
; INFORMATION FOR SEQ ID NO: 66:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 510 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; ORGANISM: Eustoma russellianum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 92 to 1621  
; IDENTIFICATION METHOD: by experiment  
; SEQUENCE DESCRIPTION: SEQ ID NO: 66  
US-09-616-990-66

Query Match 90.6%; Score 29; DB 4; Length 510;  
Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGSVLL 7  
|:||||  
DB 3 VGNVLL 9

## RESULT 3

US-08-978-458-4  
; Sequence 4, Application US/08978458  
; Patent No. 5932701  
; GENERAL INFORMATION:  
; APPLICANT: Palmer, Leslie M.  
; APPLICANT: Fedon, Jason C.  
; APPLICANT: Warren, Richard L.  
; APPLICANT: Kosmatka, Anna L.  
; APPLICANT: Shilling, Lisa K.  
; APPLICANT: Stodola, Robert K.  
; APPLICANT: Knowles, David J. C.  
; APPLICANT: Black, Michael E.  
; APPLICANT: Hodgson, John E.  
; APPLICANT: Nicholas, Richard O.  
; TITLE OF INVENTION: rida  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert, Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2793  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,458  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 08/911,503  
; FILING DATE: 15-AUG-1997  
; APPLICATION NUMBER: PCT/US97/14436  
; FILING DATE: 15-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dickinson, O. Todd  
; REGISTRATION NUMBER: 28,354  
; REFERENCE/DOCKET NUMBER: P50533-09  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215/994-2252  
; TELEFAX: 215/994-2222  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 163 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-978-458-4

Query Match 87.5%; Score 28; DB 2; Length 163;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GSGVLL 7  
|:||||  
DB 46 GSGVLL 51

## RESULT 4

US-08-978-458-2  
; Sequence 2, Application US/08978458  
; Patent No. 5932701  
; GENERAL INFORMATION:  
; APPLICANT: Palmer, Leslie M.  
; APPLICANT: Fedon, Jason C.



APPLICANT: Warren, Richard L.  
APPLICANT: Kosmatka, Anna L.  
APPLICANT: Shilling, Lisa K.  
APPLICANT: Stodola, Robert K.  
APPLICANT: Knowles, David J. C.  
APPLICANT: Black, Michael T.  
APPLICANT: Hodgson, John E.  
APPLICANT: Nicholas, Richard O.  
TITLE OF INVENTION: r1ba  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/978,458  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/911,503  
FILING DATE: 15-AUG-1997  
APPLICATION NUMBER: PCT/US97/14436  
FILING DATE: 15-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, O. Todd  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: P50533-09  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215/994-2252  
TELEFAX: 215/994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 412 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-978-458-2

Query Match 87.5%; Score 28; DB 2; Length 412;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGVLL 7  
DB 295 GSGVLL 300

RESULT 5  
US-08-745-934-3  
Sequence 3, Application US/08745934  
Patent No. 5861496  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Hawkins, Phillip R.  
TITLE OF INVENTION: HUMAN SQUALENE EPOXIDASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/745,934  
FILING DATE: Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0151 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 855-0555  
TELEFAX: (415) 845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 572 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1217593  
US-08-745-934-3

Query Match 87.5%; Score 28; DB 2; Length 572;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSVL 6  
DB 127 VSGSVL 132

RESULT 6  
US-09-147-009-6  
Sequence 6, Application US/09147009  
Patent No. 6153815  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Process for Raising Squalene Levels in Plants  
TITLE OF INVENTION: and DNA Sequences Used Therefor  
NUMBER OF SEQUENCES: 11  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/147,009  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 572 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Mus Musculus  
STRAIN: B6CBA  
DEVELOPMENTAL STAGE: 6-8 weeks  
IMMEDIATE SOURCE:  
LIBRARY: Lambda Zap vector Stratagene catalog #935302  
CLONE: pMSE-17  
PUBLICATION INFORMATION:

AUTHORS: Kosuga, K.  
AUTHORS: Hata, S.  
AUTHORS: Osumi, T.  
AUTHORS: Sakakibara, J.  
AUTHORS: Ono, T.  
TITLE: Nucleotide sequence of a cDNA for mouse  
TITLE: squalene epoxidase  
JOURNAL: Biochim. Biophys. Acta  
VOLUME: 1260  
ISSUE: 3  
PAGES: 345-348  
DATE: 1995  
US-09-147-009-6

Query Match 87.5%; Score 28; DB 4; Length 572;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSGVL 6  
DB 127 GSGVL 132

RESULT 7  
US-08-987-466-4  
; Sequence 4, Application US/08987466  
; Patent No. 5922595  
; GENERAL INFORMATION:  
; APPLICANT: Fisher, Douglas A.  
; APPLICANT: Gooding, Doug  
; APPLICANT: Streeter, Dave  
; TITLE OF INVENTION: CYCLIC-GMP PHOSPHODIESTERASE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/987,466  
; FILING DATE: Filed Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0442 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 584 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 829179  
; US-08-987-466-4

Query Match 87.5%; Score 28; DB 2; Length 584;

Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGVL 7  
DB 381 GSGVL 386

RESULT 8  
US-09-240-359-4  
; Sequence 4, Application US/09240359  
; Patent No. 6255456  
; GENERAL INFORMATION:  
; APPLICANT: Fisher, Douglas A.  
; APPLICANT: Gooding, Doug  
; APPLICANT: Streeter, Dave  
; TITLE OF INVENTION: CYCLIC-GMP PHOSPHODIESTERASE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/240,359  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/987,466  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0442 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 584 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 829179  
; US-09-240-359-4

Query Match 87.5%; Score 28; DB 4; Length 584;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSGVL 7  
DB 381 GSGVL 386

RESULT 9  
US-08-829-026A-6  
; Sequence 6, Application US/08829026A  
; Patent No. 5837825  
; GENERAL INFORMATION:  
; APPLICANT: Meinersmann, Richard J.  
; APPLICANT: Khoury, Christian A.  
; TITLE OF INVENTION: Campylobacter Jejuni Flagellin-Escherichia Coli LT-B Fusion

NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Janelle S. Graeter  
STREET: Room 411, Building 005, BARC-W  
CITY: Beltsville  
STATE: MD  
COUNTRY: USA  
ZIP: 20705  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/829,026A  
FILING DATE: 18-AUG-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Graeter, Janelle S.  
REGISTRATION NUMBER: 35,024  
REFERENCE/DOCKET NUMBER: 0106.97  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-504-5676  
TELEFAX: 301-504-5060  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 371 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-829-026A-6

Query Match 84.4%; Score 27; DB 2; Length 371;  
Best Local Similarity 83.3%; Pred. No. 5.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVL 6  
:|||||  
Db 283 VSGSIL 288

RESULT 10  
US-08-362-512A-2  
Sequence 2, Application US/08362512A  
Patent No. 5719043  
GENERAL INFORMATION:  
APPLICANT: FROMMER, Wolf-Bernd  
TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID  
TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS  
TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
STREET: 1180 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10036-8403  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,512A  
FILING DATE: 05-JAN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/007,636  
FILING DATE: 21-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mellman, Edward A.

REGISTRATION NUMBER: 24,735  
REFERENCE/DOCKET NUMBER: P/951-107  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 382-0700  
TELEFAX: (212) 382-0888  
TELEX: 236925  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 485 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-362-512A-2

Query Match 84.4%; Score 27; DB 1; Length 485;  
Best Local Similarity 83.3%; Pred. No. 7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVL 6  
:|||||  
Db 53 IGSGLV 58

RESULT 11  
US-08-964-939-2  
Sequence 2, Application US/08964939  
Patent No. 6245970  
GENERAL INFORMATION:  
APPLICANT: FROMMER, Wolf-Bernd  
TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID  
TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS  
TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
STREET: 1180 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10036-8403  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/964,939  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,512  
FILING DATE: 05-JAN-1995  
APPLICATION NUMBER: US 08/007,636  
FILING DATE: 21-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mellman, Edward A.  
REGISTRATION NUMBER: 24,735  
REFERENCE/DOCKET NUMBER: P/951-107  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 382-0700  
TELEFAX: (212) 382-0888  
TELEX: 236925  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 485 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-964-939-2

Query Match 84.4%; Score 27; DB 4; Length 485;

Best Local Similarity 83.3%; Pred. No. 7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVL 6  
:|||||  
Db 53 IGSGVL 58

RESULT 12  
US-08-362-512A-4  
; Sequence 4, Application US/08362512A  
; Patent No. 5719043  
; GENERAL INFORMATION:  
; APPLICANT: FROMMER, Wolf-Bernd  
; TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID  
; TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS  
; TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ostroienk, Faber, Gerb & Soffen  
; STREET: 1180 Avenue of the Americas  
; City: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10036-8403  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,512A  
; FILING DATE: 05-JAN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/007,636  
; FILING DATE: 21-JAN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meliman, Edward A.  
; REGISTRATION NUMBER: 24,735  
; REFERENCE/DOCKET NUMBER: P/951-107  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 382-0700  
; TELEFAX: (212) 382-0888  
; TELER: 236925  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 493 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-362-512A-4

Query Match 84.4%; Score 27; DB 1; Length 493;  
Best Local Similarity 83.3%; Pred. No. 7.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVL 6  
:|||||  
Db 62 IGSGVL 67

RESULT 13  
US-08-964-939-4  
; Sequence 4, Application US/08964939  
; Patent No. 6245970  
; GENERAL INFORMATION:  
; APPLICANT: FROMMER, Wolf-Bernd  
; TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID  
; TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS  
; TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE  
; NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ostroienk, Faber, Gerb & Soffen  
; STREET: 1180 Avenue of the Americas  
; City: New York  
; STATE: NY  
; COUNTRY: US  
; ZIP: 10036-8403  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/964,939  
; FILING DATE: 05-JAN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,512  
; FILING DATE: 05-JAN-1995  
; APPLICATION NUMBER: US 08/007,636  
; FILING DATE: 21-JAN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meliman, Edward A.  
; REGISTRATION NUMBER: 24,735  
; REFERENCE/DOCKET NUMBER: P/951-107  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 382-0700  
; TELEFAX: (212) 382-0888  
; TELER: 236925  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 493 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-964-939-4

Query Match 84.4%; Score 27; DB 4; Length 493;  
Best Local Similarity 83.3%; Pred. No. 7.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVL 6  
:|||||  
Db 62 IGSGVL 67

RESULT 14  
US-08-745-934-4  
; Sequence 4, Application US/08745934  
; Patent No. 5861496  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Hawkins, Phillip R.  
; TITLE OF INVENTION: HUMAN SQUALENE EPOXIDASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; City: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/745,934  
; FILING DATE: Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0151 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 855-0555  
TELEFAX: (415) 845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 573 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 1083804  
US-08-745-934-4

Query Match 84.4%; Score 27; DB 2; Length 573;  
Best Local Similarity 83.3%; Pred. No. 8.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGVL 6  
DB 128 IGSGVL 133

RESULT 15  
US-09-147-009-7  
Sequence 7, Application US/09147009  
Patent No. 6153815  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Process for Raising Squalene Levels in Plants  
TITLE OF INVENTION: and DNA Sequences Used Therefor  
NUMBER OF SEQUENCES: 11  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/147,009  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 573 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Rattus norvegicus  
TISSUE TYPE: kidney  
CELL LINE: NRK  
IMMEDIATE SOURCE:  
LIBRARY: pCD2 library of H. Okayama  
CLONE: Tb-1  
PUBLICATION INFORMATION:  
AUTHORS: Sakakibara, J.  
AUTHORS: Watanabe, R.  
AUTHORS: Kanao, R.  
AUTHORS: Ono, T.  
TITLE: Molecular cloning and expression of rat  
JOURNAL: J Biol. Chem.  
VOLUME: 270  
ISSUE: 1

PAGES: 17-20  
DATE: 1995  
US-09-147-009-7

Query Match 84.4%; Score 27; DB 4; Length 573;  
Best Local Similarity 83.3%; Pred. No. 8.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VSGVL 6  
DB 128 IGSGVL 133

Search completed: September 6, 2001, 16:39:38  
Job time: 127 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: September 6, 2001, 16:43:30 : Search time 225.25 Seconds  
(without alignments)  
1.884 Million cell updates/sec

Title: US-09-603-713-25

Perfect score: 32

Sequence: 1 VSGSVLL 7

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

## Database :

A.Geneseq\_0601.\*  
1: /SIDSB/gcgdata/geneseq/geneseqp/AA1960.DAT:\*  
2: /SIDSB/gcgdata/geneseq/geneseqp/AA1961.DAT:\*  
3: /SIDSB/gcgdata/geneseq/geneseqp/AA1962.DAT:\*  
4: /SIDSB/gcgdata/geneseq/geneseqp/AA1963.DAT:\*  
5: /SIDSB/gcgdata/geneseq/geneseqp/AA1964.DAT:\*  
6: /SIDSB/gcgdata/geneseq/geneseqp/AA1965.DAT:\*  
7: /SIDSB/gcgdata/geneseq/geneseqp/AA1966.DAT:\*  
8: /SIDSB/gcgdata/geneseq/geneseqp/AA1967.DAT:\*  
9: /SIDSB/gcgdata/geneseq/geneseqp/AA1968.DAT:\*  
10: /SIDSB/gcgdata/geneseq/geneseqp/AA1969.DAT:\*  
11: /SIDSB/gcgdata/geneseq/geneseqp/AA1970.DAT:\*  
12: /SIDSB/gcgdata/geneseq/geneseqp/AA1971.DAT:\*  
13: /SIDSB/gcgdata/geneseq/geneseqp/AA1972.DAT:\*  
14: /SIDSB/gcgdata/geneseq/geneseqp/AA1973.DAT:\*  
15: /SIDSB/gcgdata/geneseq/geneseqp/AA1974.DAT:\*  
16: /SIDSB/gcgdata/geneseq/geneseqp/AA1975.DAT:\*  
17: /SIDSB/gcgdata/geneseq/geneseqp/AA1976.DAT:\*  
18: /SIDSB/gcgdata/geneseq/geneseqp/AA1977.DAT:\*  
19: /SIDSB/gcgdata/geneseq/geneseqp/AA1978.DAT:\*  
20: /SIDSB/gcgdata/geneseq/geneseqp/AA1979.DAT:\*  
21: /SIDSB/gcgdata/geneseq/geneseqp/AA2000.DAT:\*  
22: /SIDSB/gcgdata/geneseq/geneseqp/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	100.0	7	22	AA66583
2	32	100.0	7	22	AA661345
3	32	100.0	10	22	AA66588
4	32	100.0	10	22	AA661350
5	30	93.8	96	21	AA633270
6	30	93.8	103	21	AA633269
7	30	93.8	442	21	AA636904
8	30	93.8	471	21	AA636903
9	30	93.8	477	21	AA636902
10	29	90.6	140	21	AA695315
11	29	90.6	221	21	AA695316

12	29	90.6	510	14	AA640867
13	28	87.5	163	19	AA68142
14	28	87.5	163	20	AA68135
15	28	87.5	175	21	AA641215
16	28	87.5	243	20	AA682283
17	28	87.5	254	20	AA682285
18	28	87.5	324	21	AA68832
19	28	87.5	388	21	AA684116
20	28	87.5	412	20	AA618134
21	28	87.5	428	20	AA635302
22	28	87.5	545	21	AA607575
23	28	87.5	550	20	AA634443
24	28	87.5	554	20	AA634442
25	28	87.5	562	20	AA634322
26	28	87.5	709	21	AA691456
27	28	87.5	717	21	AA621231
28	28	87.5	3023	17	AA694462
29	27	84.4	11	21	AA607256
30	27	84.4	148	21	AA642601
31	27	84.4	175	21	AA644765
32	27	84.4	180	21	AA628428
33	27	84.4	198	21	AA627975
34	27	84.4	224	21	AA644769
35	27	84.4	245	21	AA654023
36	27	84.4	294	21	AA644767
37	27	84.4	303	21	AA644768
38	27	84.4	305	21	AA630828
39	27	84.4	316	21	AA627974
40	27	84.4	332	21	AA630827
41	27	84.4	336	21	AA693903
42	27	84.4	343	21	AA627973
43	27	84.4	350	21	AA617373
44	27	84.4	350	21	AA647682
45	27	84.4	352	21	AA617372

## ALIGNMENTS

RESULT 1  
ID AAB6583 standard; Peptide: 7 AA.  
AC AAB6583:  
DT 12-APR-2001 (first entry)  
DE Synthetic memapsin 2 substrate peptide #2.  
KW memapsin 2; neotropic; neuroprotective; amyloid precursor protein;  
KW APP; memapsin 2 inhibitor; Alzheimer's disease.  
OS Synthetic.  
XX WO200100665-A2.  
XX PD 04-JAN-2001.  
XX PF 27-JUN-2000; 2000MO-US17742.  
XX PR 28-JUN-1999; 99US-0141363.  
XX PR 30-NOV-1999; 99US-0168060.  
XX PR 25-JAN-2000; 2000US-0177836.  
XX PR 27-JAN-2000; 2000US-0178368.  
XX PR 08-JUN-2000; 2000US-0210292.  
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
XX (UNIT ) UNIV ILLINOIS FOUND.  
XX Tang JUN, Hong L, Ghosh AK;  
XX WPI: 2001-137933/14.

Flavonoid-3',5'-hy  
S. pneumoniae deri  
S. pneumoniae riba  
Human ORFX ORF979  
Tumour surface pro  
Natural killer cell  
Breast and ovarian  
Plytyl/Prelytrans  
S. pneumoniae riba  
Amino acid sequenc  
Protein encoded by  
Porphyromonas ging  
Porphyromonas ging  
Porphyromonas ging  
Human secreted pro  
Tomato Lemp1. Ly  
Hepatitis C virus  
Human peptidomime  
Human ORFX ORF2365  
Corn serine O-acet  
Wheat branched cha  
Arabidopsis thalia  
Rice serine O-acet  
Human pancreatic c  
Impatiens balsamia  
Rice serine O-acet  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Amino acid sequenc  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia

PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage  
 XX  
 PS Example 4; Page 33; 86pp; English.  
 XX  
 CC The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 CC  
 SO Sequence 7 AA:  
 Query Match 100.0%; Score 32; DB 22; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSVLL 7  
 Db 1 VSGSVLL 7  
 RESULT 2  
 AAB61345  
 ID AAB61345 standard; peptide; 7 AA.  
 AC AAB61345;  
 XX  
 DT 02-APR-2001 (first entry)  
 DE Memapsin 2 substrate #4.  
 XX  
 DE Memapsin 2; catalyst; Alzheimer's.  
 XX  
 KM Memapsin 2; catalyst; Alzheimer's.  
 XX  
 OS Synthetic.  
 OS  
 PN WO200100663-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 27-JUN-2000; 2000MO-US17661.  
 XX  
 PR 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX  
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 XX  
 PI Tang JUN, Lin X, Koelsch G;  
 PI  
 DR WPI; 2001-102885/11.  
 XX  
 PT Purified recombinant catalytically active memapsin 2, used to screen  
 PT inhibitors of it, which are used to treat and prevent Alzheimer's  
 PT disease -  
 XX  
 PS Example 4; Page 33; 86pp; English.  
 XX  
 CC The present invention relates to a purified recombinant  
 CC catalytically active memapsin 2. The invention may be used for  
 CC isolating inhibitors which are used to treat or prevent  
 CC Alzheimer's disease. The invention may also be used to screen  
 CC for individuals more genetically prone to develop Alzheimer's  
 CC disease.

XX  
 SO Sequence 7 AA:  
 Query Match 100.0%; Score 32; DB 22; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSVLL 7  
 Db 1 VSGSVLL 7  
 RESULT 3  
 AAB66588  
 ID AAB66588 standard; Peptide; 10 AA.  
 AC AAB66588;  
 XX  
 DT 12-APR-2001 (first entry)  
 DE Synthetic memapsin 2 substrate peptide.  
 XX  
 DE Synthetic memapsin 2 substrate peptide.  
 XX  
 KM memapsin 2; neurotropic; neuroprotective; amyloid precursor protein;  
 KM APP; memapsin 2 inhibitor; Alzheimer's disease.  
 XX  
 OS Synthetic.  
 OS  
 PN WO200100665-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 27-JUN-2000; 2000MO-US17742.  
 XX  
 PR 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX  
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA (UNIT ) UNIT ILLINOIS FOUND.  
 XX  
 PI Tang JUN, Hong L, Ghosh AK;  
 PI  
 DR WPI; 2001-137933/14.  
 XX  
 PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage -  
 XX  
 PS Example 4; Page 33; 86pp; English.  
 XX  
 CC The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 CC  
 SO Sequence 10 AA:  
 Query Match 100.0%; Score 32; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSVLL 7  
 Db 1 VSGSVLL 7



DB 1 vgsqvl1 7

RESULT 4  
AAB61350  
ID AAB61350 standard; peptide; 10 AA.  
XX  
AC AAB61350;  
XX  
DT 02-APR-2001 (first entry)  
XX  
DE Synthetic peptide #2.  
XX  
XX Memapsin 2; catalyst; Alzheimer's.  
XX  
OS Synthetic.  
XX  
PN WO200100663-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 27-JUN-2000; 2000WO-US17661.  
XX  
PR 28-JUN-1999; 99US-0141363.  
XX  
PR 30-NOV-1999; 99US-0168060.  
XX  
PR 25-JAN-2000; 2000US-0177836.  
XX  
PR 27-JAN-2000; 2000US-0176368.  
XX  
PR 08-JUN-2000; 2000US-0210292.  
XX  
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
XX  
PI Tang JUN, Lin X, Koelsch G;  
XX  
DR MPI: 2001-102885/11.  
XX  
XX Purified recombinant catalytically active memapsin 2, used to screen  
PT Inhibitors of it, which are used to treat and prevent Alzheimer's  
PT disease -  
XX  
XX  
PS Disclosure: Page 86; 86pp; English.  
XX  
XX  
CC The present invention relates to a purified recombinant  
CC catalytically active memapsin 2. The invention may be used for  
CC isolating inhibitors which are used to treat or prevent  
CC Alzheimer's disease. The invention may also be used to screen  
CC for individuals more genetically prone to develop Alzheimer's  
CC disease.  
XX  
SQ Sequence 10 AA;

Query Match 100.0%; Score 32; DB 22; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
|||||  
Db 1 vgsqvl1 7

RESULT 5  
AAB33270  
ID AAB33270 standard; Protein; 96 AA.  
XX  
AC AAB33270;  
XX  
DT 25-JAN-2001 (first entry)  
XX  
DE Pinus radiata transcription factor protein sequence #316.  
XX  
XX Plant; transcription factor; gene expression; eucalyptus; pine;  
KM poplar; sweetgum; teak; mahogany; bzif; G-box binding factor;  
KM basic helix-loop-helix zipper; homeotic; homeobox; MADS;  
KW basic helix-loop-helix zipper; homeotic; homeobox; MADS;

KW homeodomain zipper; LIM domain; AP2; EREBS; zinc finger domain;  
KW type 2 Cys2His2; CCAAT box element; MYB.  
XX  
OS Pinus radiata.  
XX  
PN WO200053724-A2.  
XX  
PD 14-SEP-2000.  
XX  
PF 09-MAR-2000; 2000WO-US06112.  
XX  
PR 11-MAR-1999; 99US-0266513.  
XX  
PR 18-AUG-1999; 99US-0149485.  
XX  
PA (GENE-) GENESIS RES & DEV CORP LTD.  
XX  
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
XX  
PI Wood M, McGrath A, Shenk MA, Glenn M;  
XX  
PI Wood M, McGrath A, Shenk MA, Glenn M;  
XX  
DR MPI: 2000-579369/54.  
XX  
XX  
PT New isolated polynucleotide encoding a plant transcription factor for  
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
PT having modified gene expression or modified activity of a polypeptide  
XX  
XX  
PS Claim 8; Page 702; 747pp; English.  
XX  
XX  
CC The present invention relates to novel plant transcription factors from  
CC Eucalyptus grandis or Pinus radiata. The present sequence is one such  
CC transcription factor. The transcription factor may be used to produce a  
CC plant having modified gene expression such as a woody plant e.g. a  
CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
CC to modify the activity of a polypeptide in a plant. The transcription  
CC factors of the present invention are members from the following families  
CC of regulatory proteins: bzif, bzif family of G-box binding factors, basic  
CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
CC zipper, LIM domain, AP2 and EREBS, zinc finger domains of type 2  
CC Cys2His2, CCAAT box elements and MYB.  
XX  
SQ Sequence 96 AA;

Query Match 93.8%; Score 30; DB 21; Length 96;  
Best Local Similarity 85.7%; Pred. No. 46;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
|||||  
Db 20 vgsqvl1 26

RESULT 6  
AAB33269  
ID AAB33269 standard; Protein; 103 AA.  
XX  
AC AAB33269;  
XX  
DT 25-JAN-2001 (first entry)  
XX  
DE Pinus radiata transcription factor protein sequence #315.  
XX  
XX Plant; transcription factor; gene expression; eucalyptus; pine;  
KM poplar; sweetgum; teak; mahogany; bzif; G-box binding factor;  
KM basic helix-loop-helix zipper; homeotic; homeodomain; homeobox;  
KM homeodomain zipper; LIM domain; AP2; EREBS; zinc finger domain;  
KW type 2 Cys2His2; CCAAT box element; MYB.  
XX  
XX Pinus radiata.  
XX  
OS  
XX  
PN WO200053724-A2.  
XX  
PD 14-SEP-2000.

XX 09-MAR-2000; 2000MO-US06112.  
PF 11-MAR-1999; 99US-0266513.  
PR 18-AUG-1999; 99US-0149485.  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
XX Wood M, McGrath A, Shank MA, Glenn M;  
PI WPI; 2000-579369/54.  
DR  
XX  
XX New isolated polynucleotide encoding a plant transcription factor for  
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
PT having modified gene expression or modified activity of a polypeptide  
PT  
XX  
XX Claim 8; Page 702; 747pp; English.  
PS  
XX The present invention relates to novel plant transcription factors from  
CC Eucalyptus grandis or Pinus radiata. The present sequence is one such  
CC transcription factor. The transcription factor may be used to produce a  
CC plant having modified gene expression such as a woody plant e.g. a  
CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
CC to modify the activity of a polypeptide in a plant. The transcription  
CC factors of the present invention are members from the following families  
CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
CC zipper, LIM domain, AP2 and ERFs, zinc finger domains of type 2  
CC Cys2His2, CCAAT box elements and MYB.  
XX  
SQ Sequence 103 AA;

Query Match 93.8%; Score 30; DB 21; Length 103;  
Best Local Similarity 85.7%; Pred. No. 50;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSVLL 7  
|||!|!  
Db 20 vsgsvml 26

RESULT 7  
ID AAG36904 standard; Protein: 442 AA.  
XX AAG36904;  
XX  
XX 18-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45291.  
XX  
XX Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX EP1033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140921.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 93.8%; Score 30; DB 21; Length 442;  
Best Local Similarity 85.7%; Pred. No. 2.4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VCSGVL 7  
||||:|  
DB 346 vsgsvll 352

RESULT 8  
ID AAG36903 standard; Protein: 471 AA.  
XX AAG36903;  
AC 18-OCT-2000 (first entry)  
XX  
DT  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45290.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0123825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147312.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 93.8%; Score 30; DB 21; Length 471;  
Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGSGLVLL 7  
|||||:  
DB 375 vgsavl 381

RESULT 9  
AAG36902  
ID AAG36902 standard; Protein; 477 AA.

XX AAG36902;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 45289.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hydrolisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX OS EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-012180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134229.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143542.  
PR 13-JUL-1999; 99US-0143547.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 10-AUG-1999; 99US-0148319.  
 PR 11-AUG-1999; 99US-0148341.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157553.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 14-OCT-1999; 99US-0159684.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 22-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 93.8%; Score 30; DB 21; Length 477;  
 Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGCVLL 7  
 DB 381 VSGSVLL 387

## RESULT 10

AAV95315 standard; Protein; 140 AA.

AAV95315;

12-SEP-2000 (first entry)

Soybean phosphatidylyllycerophosphate synthase.

Soybean phosphatidylyllycerophosphate synthase; phospholipid;  
 transgenic plant; herbicide.

Glycine max.

WO200036117-A1.

22-JUN-2000.

15-DEC-1999; 99WO-US29826.

16-DEC-1998; 98US-0112558.

(DUPO ) DU PONT DE NEMOURS & CO E I.

Cahoon RE, Falco SC, Kinney AJ;

MP1: 2000-431595/37.  
 N-PSDB; AAA27932.

Nucleic acids encoding plant GDP (cytosine diphosphate)-alcohol  
 phosphatidyltransferase polypeptide, useful for creating transgenic  
 plants in which the polypeptides are present at higher or lower levels  
 than normal -

Claim 20; page 47-48; 59pp; English.

PS The present sequence is that of soybean phosphatidylyllycerophosphate  
 synthase, an enzyme that catalyzes the formation of  
 phosphatidylyllycerophosphate, the only phospholipid in chloroplast  
 thylakoids. The sequence was deduced from a contig of isolated  
 cDNA clones (see AAA27932). The invention provides plant phospholipid  
 biosynthetic enzymes (see AAV95311-20) and the nucleic acids (see  
 AAA27928-37). The nucleic acids are useful for creating chimeric  
 genes encoding all or a portion of the phospholipid biosynthetic  
 enzyme, in sense or antisense orientation, where expression of the  
 chimeric gene results in production of altered levels of the enzyme  
 in a transformed host cell. Transgenic plants express the enzymes  
 at higher or lower levels than normal, or in cell types or  
 developmental stages in which they are not normally found.

CC Phospholipid biosynthetic enzymes expressed by host cells can be  
 CC used to identify inhibitors that may be useful as herbicides.

XX Sequence 140 AA:

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSVLL 7  
 |||||:  
 Db 190 vsgsvll 196

Query Match 90.6%; Score 29; DB 21; Length 140;  
 Best Local Similarity 85.7%; Pred. No. 1.1e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 |||||:  
 Db 109 vsgsvll 115

RESULT 11

AA95316 ID AA95316 standard; Protein; 221 AA.

AC AAY95316;

DT 12-SEP-2000 (first entry)

XX Soybean phosphatidylglycerophosphate synthase.

KM Soybean: phosphatidylglycerophosphate synthase; phospholipid;

transgenic plant; herbicide.

OS Glycine max.

XX MO200036117-A1.

PN 22-JUN-2000.

PD 15-DEC-1999; 99MO-US29826.

XX 16-DEC-1998; 980S-0112558.

PR (DUPO ) DU PONT DE NEMOURS & CO E I.

PA Cahoon RE, Falco SC, Kinney AJ;

PI WPI: 2000-431595/37.

DR N-PSDB; AA27933.

XX Nucleic acids encoding plant GDP (cytosine diphosphate)-alcohol  
 PT phosphatidyltransferase polypeptide, useful for creating transgenic  
 PT plants in which the polypeptides are present at higher or lower levels  
 PT than normal -  
 PS Claim 20; Page 48-49; 59pp; English.

XX The present sequence is that of a C-terminal portion of soybean  
 CC phosphatidylglycerophosphate synthase, an enzyme that catalyzes the  
 CC formation of phosphatidylglycerophosphate, the only phospholipid in  
 CC thylakoids. The sequence was deduced from a contig of isolated  
 CC cDNA clones (see AA27933). The invention provides plant phospholipid  
 CC biosynthetic enzymes (see AAY95311-20) and the nucleic acids (see  
 CC AA27928-37). The nucleic acids are useful for creating chimeric  
 CC genes encoding all or a portion of the phospholipid biosynthetic  
 CC enzyme, in sense or antisense orientation, where expression of the  
 CC chimeric gene results in production of altered levels of the enzyme  
 CC in a transformed host cell. Transgenic plants express the enzymes  
 CC at higher or lower levels than normal, or in cell types or  
 CC developmental stages in which they are not normally found.  
 CC Phospholipid biosynthetic enzymes expressed by host cells can be  
 CC used to identify inhibitors that may be useful as herbicides.

XX Sequence 221 AA:

Query Match 90.6%; Score 29; DB 21; Length 221;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 |||||:  
 Db 190 vsgsvll 196

RESULT 12

AA40867 ID AA40867 standard; Protein; 510 AA.

AC AA40867;

DT 28-MAR-1994 (first entry)

XX Flavonoid-3',5'-hydroxylase.

KM Flavonoid-3',5'-hydroxylase; transformation; plants; petunia; rose;

tobacco; pigment alteration; blue; SFP; single specific primer.

OS Eustoma russellianum.

XX MO9318155-A.

PN 16-SEP-1993.

PD 20-NOV-1992; 92WO-JP01520.

XX 02-MAR-1992; 92JP-0044963.

PR (KYOW ) KYOWA HAKKO KOGYO CO LTD.

PA Kikuchi Y, Kiyokawa S, Ohbayashi M, Okinaka Y, Shimada R;

PI Shimada Y;

DR WPI: 1993-303469/38.

DR N-PSDB; AA047840.

XX Gene coding for flavonoid-3',5'-hydroxylase of petunia petals -  
 PT used to transform plants e.g. petunia, rose or tobacco to give  
 PT bluer flower colour and altered pigment pattern  
 PS Claim 1; Page 67-71; 82pp; Japanese.

XX Insertion of the sequences (AA047840-42) into plants such as rose,  
 CC petunia, tobacco and carnation, using a suitable vector such as  
 CC agrobacterium, give transformed plants which express the gene,  
 CC resulting in petals with a bluer colour than normal, and/or  
 CC pigmentation patterns which do not occur naturally. The sequences  
 CC were amplified using primers (AA047843-70). Related single specific  
 CC primers using a gene sequence coding for the haem-binding region of  
 CC cytochrome P450 are shown in (AA047871-047903).

XX Sequence 510 AA:

Query Match 90.6%; Score 29; DB 14; Length 510;  
 Best Local Similarity 85.7%; Pred. No. 4.4e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSVLL 7  
 |||||:  
 Db 3 vngsvll 9

RESULT 13

AA86142 ID AAY86142 standard; Protein; 163 AA.

AC AAY86142;

DT 10-APR-2000 (first entry)





```

XX      WO200058473-A2.
PN
XX
XX      05-OCT-2000.
PD
XX
XX      31-MAR-2000; 2000WO-US08621.
PF
XX
XX      31-MAR-1999; 99US-0127607.
PR      02-APR-1999; 99US-0127636.
PR      05-APR-1999; 99US-0127728.
PR      30-MAR-2000; 2000US-0540763.
XX
XX      (CURA-) CURAGEN CORP.
PA
XX
XX      Shlunkers RA, Leach M;
PI
XX
XX      WPI; 2000-602362/57.
DR      N-PSDB; AAC75424.
XX
XX      Novel nucleic acids and peptides derived from open reading frame X,
PT      useful for treating e.g. cancers, proliferative disorders,
PT      neurodegenerative disorders and cardiovascular disease -
XX
XX      Claim 11; Page 1484; 5507pp; English.
PS
XX
XX      AAC74446 to AAC77606 encode the proteins given in AA340237 to AAB43397,
CC      which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC      sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC      antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC      osteopathic; anticonvulsant; antidiabetic; immunosuppressant;
CC      immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC      antidiabetic; hypotensive; dermatologic; immunosuppressive;
CC      antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
CC      antihypertensive; and antianemic. The sequences can be used for determining
CC      the presence of or predisposition to, or preventing or treating
CC      pathological conditions associated with an ORFX-associated disorder. The
CC      nucleic acids can be used to express ORFX proteins in gene therapy
CC      vectors. The proteins and nucleic acids may be used to treat cancers,
CC      proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC      graft vs host disease, cardiovascular disease, diabetes mellitus,
CC      hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC      erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC      bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC      allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC      nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC      coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX      Sequence 175 AA:
SQ

```

```

Query Match      87.5%; Score 28; DB 21; Length 175;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      2 GSGVLL 7
       111111
DB      158 gsgvll 163

```

Search completed: September 6, 2001, 16:43:31  
Job time: 360 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:49:52 ; Search time 231.42 Seconds  
(without alignments)  
2.859 Million cell updates/sec

Title: US-09-603-713-24

Perfect score: 24

Sequence: 1 VGS GV 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_16:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mmc:\*

8: sp\_organella:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rudent:\*

12: sp\_unclassified:\*

13: sp\_vertebrate:\*

14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	24	100.0	88	5	Q20658	Q20658 caenorhabd
2	24	100.0	91	2	P77018	P77018 escherichia
3	24	100.0	113	2	Q9Z6U8	Q9Z6U8 chlamydia p
4	24	100.0	113	2	Q9JRZ4	Q9JRZ4 chlamydia p
5	24	100.0	122	2	Q49710	Q49710 mycobacteri
6	24	100.0	123	4	Q9P1N7	Q9P1N7 homo sapien
7	24	100.0	131	13	Q9PV19	Q9PV19 harpaxon sp
8	24	100.0	136	2	Q56518	Q56518 unidentified
9	24	100.0	136	2	Q56525	Q56525 unidentified
10	24	100.0	136	2	Q56526	Q56526 unidentified
11	24	100.0	136	2	Q56528	Q56528 unidentified
12	24	100.0	138	2	Q9XBV4	Q9XBV4 mycobacteri
13	24	100.0	146	1	Q9YTAB9	Q9YTAB9 aeropyrum p
14	24	100.0	149	4	Q9RYV8	Q9RYV8 deinococcus
15	24	100.0	150	4	Q9Y4T1	Q9Y4T1 homo sapien
16	24	100.0	165	1	Q9HNT3	Q9HNT3 halobacteri
17	24	100.0	180	10	Q64541	Q64541 arabidopsis
18	24	100.0	182	14	Q85656	Q85656 moloney mur
19	24	100.0	187	5	Q9N9N2	Q9N9N2 leishmania

20	24	100.0	196	10	082245	082245 arabidopsis
21	24	100.0	200	2	Q48974	Q48974 mycoplasma
22	24	100.0	201	2	Q48808	Q48808 lactobacilli
23	24	100.0	207	2	Q45491	Q45491 bacillus st
24	24	100.0	217	2	Q53373	Q53373 synechococc
25	24	100.0	217	10	Q9FE82	Q9FE82 arabidopsis
26	24	100.0	217	14	Q91IH6	Q91IH6 frog adenov
27	24	100.0	226	5	Q9W455	Q9W455 drosophila
28	24	100.0	227	2	Q9K7U1	Q9K7U1 bacillus ha
29	24	100.0	229	10	Q9LUF9	Q9LUF9 arabidopsis
30	24	100.0	239	11	Q9WV68	Q9WV68 mus musculu
31	24	100.0	239	11	P70455	P70455 mus musculu
32	24	100.0	251	2	Q32918	Q32918 mus musculu
33	24	100.0	253	2	Q9EZM3	Q9EZM3 isophaera
34	24	100.0	267	1	Q9HMJ2	Q9HMJ2 streptococci
35	24	100.0	278	2	Q9RC36	Q9RC36 streptococci
36	24	100.0	282	5	Q94513	Q94513 drosophila
37	24	100.0	283	5	Q17439	Q17439 caenorhabd
38	24	100.0	283	5	Q24260	Q24260 drosophila
39	24	100.0	284	2	Q9F2W8	Q9F2W8 streptococci
40	24	100.0	287	1	Q9Y937	Q9Y937 aeropyrum p
41	24	100.0	291	1	Q9HKW7	Q9HKW7 thermoplasma
42	24	100.0	306	5	Q19660	Q19660 caenorhabd
43	24	100.0	309	10	Q80448	Q80448 arabidopsis
44	24	100.0	310	2	Q24886	Q24886 heliobacte
45	24	100.0	310	2	Q84974	Q84974 heliobacte

#### ALIGNMENTS

RESULT 1

ID 020658 PRELIMINARY: PRT: 88 AA.

AC Q20658: 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

DE COSMID F52E1.

GN F52E1.6.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., Gardner A., Green P., Hawkins J., Kirsten J., Laister N., Latreille P., Jones M., Kershaw J., Murray A., Mortimore B., O'Callaghan M., Parsons J., Percy C., Ritken L., Koopa A., Saunders D., Shownkeen R., Smaiden N., Smith A., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.; \*2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.\*;

RT Nature 368:32-38(1994).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA le T.T., Rifkin L.;

RL Submitted (Dec-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Waterston R.;

RL Submitted (Nov-1995) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: U41109; AAB37040.1; -  
 SQ SEQUENCE 88 AA; 9586 MW; 119E463247ECF17D CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 24; DB 5; Length 88;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
 |||||  
 DB 70 VSGSV 74

RESULT 2  
 ID P77018 PRELIMINARY; PRT; 91 AA.  
 AC P77018;  
 DT 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)  
 DE PERIOD CLOCK PROTEIN (FRAGMENT).  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-97349980; PubMed-9205837;  
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsushashi N.,  
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,  
 RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasundaram S.,  
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Wada C.,  
 RA Yamagata S., Horiiuchi T.;  
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
 RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and  
 RT analysis of its sequence features.";  
 RL DNA Res. 4:91-113(1997).  
 DR EMBL: D90889; BAA16515.1; -  
 SQ SEQUENCE 91 AA; 9089 MW; 71085EFA9CD1FC8 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 24; DB 2; Length 91;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
 |||||  
 DB 41 VSGSV 45

RESULT 3  
 ID 0926U8 PRELIMINARY; PRT; 113 AA.  
 AC 0926U8;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE CT809 HYPOTHETICAL PROTEIN.  
 GN CPN0960.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CWL029;  
 RX MEDLINE-99206606; PubMed-10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lamell C., Fan J., Hyman R.W.,  
 RA Olyer L., Grimwood J., Davis R.W., Stephens R.S.;

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 DR EMBL: AE001676; AAD19096.1; -  
 SQ SEQUENCE 113 AA; 12789 MW; 9BF952B3968BA45A CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 24; DB 2; Length 113;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
 |||||  
 DB 58 VSGSV 62

RESULT 4  
 ID 09JR24 PRELIMINARY; PRT; 113 AA.  
 AC 09JR24;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE CT809 HYPOTHETICAL PROTEIN.  
 GN CPJ0960 OR CPJ0900.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C138;  
 RX MEDLINE-20330349; PubMed-10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE-20150255; PubMed-10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,  
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,  
 RA Bowman C., Dodson R., Gwin M., Nelson W., DeBoy R., Kolonay J.,  
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AF002548; BAA99168.1; -  
 DR EMBL: AE002249; AAF38686.1; -  
 DR TIGR: CP0900; -  
 SQ SEQUENCE 113 AA; 12816 MW; 9BF952B3969A545A CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 24; DB 2; Length 113;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
 |||||  
 DB 58 VSGSV 62

RESULT 5  
 ID 049710 PRELIMINARY; PRT; 122 AA.  
 AC 049710;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last annotation update)  
 DE COSMID B1549.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Smith D.R.;  
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U00014; AAA50887.1; -;  
 SQ SEQUENCE 122 AA; 12850 MW; 9967E3A0FA9CEFS CRC64;

Query Match 100.0%; Score 24; DB 2; Length 122;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
 |||||  
 DB 34 VSGSV 38

RESULT 6  
 O9P1N7 PRELIMINARY; PRT: 123 AA.  
 AC O9P1N7:  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DE PRO0974.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RA Zhang C., Yu Y., Zhang S., Ouyang S., Luo L., Wei H., Zhou G.,  
 RA Zhou W., Bi J., Zhang Y., Liu M., He F.;  
 RT "Functional prediction of the coding sequences of 32 new genes deduced  
 RT by analysis of cDNA clones from human fetal liver."  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF11685; AAF29584.1; -;  
 SQ SEQUENCE 123 AA; 13607 MW; 98C114676AE2FD0E CRC64;

Query Match 100.0%; Score 24; DB 4; Length 123;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
 |||||  
 DB 35 VSGSV 39

RESULT 7  
 O9PVI9 PRELIMINARY; PRT: 131 AA.  
 AC O9PVI9:  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE RAG1 PROTEIN (FRAGMENT).  
 GN RAG1.

OS Harpdon sp.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Aulopiformes; Synodontidae; Harpdoninae; Harpdon.  
 OX NCBI\_TaxID=94310;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MHCIIBDCK;  
 RX MEDLINE=9398697; PubMed=10468597;  
 RA Venkatesh B., Ning Y., Brenner S.;  
 RT "Late changes in spliceosomal introns define clades in vertebrate  
 RT evolution."  
 RT Proc. Natl. Acad. Sci. U.S.A. 96:10267-10271(1999).  
 RL EMBL: AF137178; AAD54548.1; -;  
 FT NON\_TER 1 131  
 FT NON\_TER 1 131  
 SQ SEQUENCE 131 AA; 14436 MW; 0470189338608B8 CRC64;

Query Match 100.0%; Score 24; DB 13; Length 131;  
 Best Local Similarity 100.0%; Pred. No. 5.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
 |||||  
 DB 44 VSGSV 48

RESULT 8  
 O36518 PRELIMINARY; PRT: 136 AA.  
 AC O36518:  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE ALPHA SUBUNIT OF DINITROGENASE (MOFE PROTEIN) (FRAGMENT).  
 GN NIFD.  
 OS unidentified nitrogen-fixing bacteria.  
 OC Bacteria.  
 OX NCBI\_TaxID=34107;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ueda T.;  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95254398; PubMed=7736355;  
 RA Ueda T., Suga Y., Yahiro N., Matsuguchi T.;  
 RT "Genetic diversity of N2-fixing bacteria associated with rice roots by  
 RT molecular evolutionary analysis of a nifD library."  
 RL Can. J. Microbiol. 41:235-240(1995).  
 DR EMBL: D26271; BAA05333.1; -;  
 DR HSP: P07328; 3MIN.  
 DR InterPro: IPR000510; -;  
 DR Pfam: PF00148; oxidored\_nitro; 1.  
 FT NON\_TER 1 136  
 FT NON\_TER 1 136  
 SQ SEQUENCE 136 AA; 15412 MW; D2BD0542BD6BD0FA CRC64;

Query Match 100.0%; Score 24; DB 2; Length 136;  
 Best Local Similarity 100.0%; Pred. No. 6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
 |||||  
 DB 120 VSGSV 124

RESULT 9  
 O56525 PRELIMINARY; PRT: 136 AA.  
 AC O56525:  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE RAG1 PROTEIN (FRAGMENT).  
 GN RAG1.

DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE ALPHA SUBUNIT OF DINITROGENASE (MOFE PROTEIN) (FRAGMENT).  
 GN NIFD.  
 OS unidentified nitrogen-fixing bacteria.  
 OC Bacteria.  
 OX NCB1\_TaxID=34107;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ueda T.;  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95254398; PubMed=7736355;  
 RA Ueda T., Suga Y., Yahiro N., Matsuguchi T.;  
 RT "Genetic diversity of N2-fixing bacteria associated with rice roots by  
 molecular evolutionary analysis of a nifd library.";  
 RL Can. J. Microbiol. 41:235-240(1995).  
 DR EMBL; D26274; BAA05336.1; -.  
 DR HSSP; P07328; 3MIN.  
 DR InterPro: IPR000510; -.  
 DR Pfam: PF00148; oxidored\_nitro; 1.  
 FT NON\_TER 1  
 FT SEQUENCE 136 AA; 15439 MW; 499F2223ACE52791 CRC64;  
 SO

Query Match 100.0%; Score 24; DB 2; Length 136;  
 Best Local Similarity 100.0%; Pred. No. 6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGSV 5  
 Db 120 VSGSV 124

RESULT 10  
 ID Q56526 PRELIMINARY; PRT; 136 AA.  
 AC Q56526;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE ALPHA SUBUNIT OF DINITROGENASE (MOFE PROTEIN) (FRAGMENT).  
 GN NIFD.  
 OS unidentified nitrogen-fixing bacteria.  
 OC Bacteria.  
 OX NCB1\_TaxID=34107;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ueda T.;  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95254398; PubMed=7736355;  
 RA Ueda T., Suga Y., Yahiro N., Matsuguchi T.;  
 RT "Genetic diversity of N2-fixing bacteria associated with rice roots by  
 molecular evolutionary analysis of a nifd library.";  
 RL Can. J. Microbiol. 41:235-240(1995).  
 DR EMBL; D26275; BAA05337.1; -.  
 DR HSSP; P07328; 3MIN.  
 DR InterPro: IPR000510; -.  
 DR Pfam: PF00148; oxidored\_nitro; 1.  
 FT NON\_TER 1  
 FT SEQUENCE 136 AA; 15347 MW; A6360F83E48E8722 CRC64;  
 SO

Query Match 100.0%; Score 24; DB 2; Length 136;  
 Best Local Similarity 100.0%; Pred. No. 6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
 Db 120 VSGSV 124  
 RESULT 11  
 ID Q56528 PRELIMINARY; PRT; 136 AA.  
 AC Q56528;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE ALPHA SUBUNIT OF DINITROGENASE (MOFE PROTEIN) (FRAGMENT).  
 GN NIFD.  
 OS unidentified nitrogen-fixing bacteria.  
 OC Bacteria.  
 OX NCB1\_TaxID=34107;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ueda T.;  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95254398; PubMed=7736355;  
 RA Ueda T., Suga Y., Yahiro N., Matsuguchi T.;  
 RT "Genetic diversity of N2-fixing bacteria associated with rice roots by  
 molecular evolutionary analysis of a nifd library.";  
 RL Can. J. Microbiol. 41:235-240(1995).  
 DR EMBL; D26277; BAA05339.1; -.  
 DR HSSP; P07328; 3MIN.  
 DR InterPro: IPR000510; -.  
 DR Pfam: PF00148; oxidored\_nitro; 1.  
 FT NON\_TER 1  
 FT SEQUENCE 136 AA; 15443 MW; 1CF51B4C5F0E5406 CRC64;  
 SO

Query Match 100.0%; Score 24; DB 2; Length 136;  
 Best Local Similarity 100.0%; Pred. No. 6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGSV 5  
 Db 120 VSGSV 124

RESULT 12  
 ID Q9XBV4 PRELIMINARY; PRT; 138 AA.  
 AC Q9XBV4;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE RIESKE IRON-SULFUR PROTEIN.  
 GN OCRA.  
 OS Mycobacterium smegmatis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacterium.  
 OX NCB1\_TaxID=1772;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC2 155;  
 RA Bardon F., Martinez R., Puech V., Bleiber G., Prod'hom G., Daffe M.,  
 Telenti A.;  
 RT "Genetic complementation of growth-attenuated Mycobacterium smegmatis  
 resistant to ethambutol.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF155062; AAD42391.1; -.  
 DR HSSP; P08980; 1RFS.  
 DR InterPro: IPR001281; -.  
 DR Pfam: PF00335; RIESKE; 1.  
 DR PRINTS; PRO0162; RIESKE.  
 DR PROSITE; PS00200; RIESKE\_2; UNKNOWN\_1.

SO SEQUENCE 138 AA: 13661 MW: 7AB6B6250812DC84 CRC64;

Query Match Best Local Similarity 100.0%; Score 24; DB 2; Length 138;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 58 VSGSV 62

RESULT 13

O9YAB9 PRELIMINARY; PRT; 146 AA.

AC O9YAB9: 01-NOV-1999 (TREMblrel. 12, Created)  
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
GN APE2020.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;  
CC Aeropyrum.  
OX NCBI\_TaxID=56636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K1.  
RX MEDLINE=99310339; PubMed=10382966;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
RA Jit-no K., Takahashi M., Sekine M., Baba S.-I., Nakai A., Kosugi H.,  
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RT crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).  
DR EMBL; AF000063; BAA81030.1; -  
DR InterPro: IPR002478; -  
DR SMART: SM00359; PUA: 1.  
KW Hypothetical protein.  
SQ SEQUENCE 146 AA: 15854 MW: AFE0629DF2F0995 CRC64;

Query Match Best Local Similarity 100.0%; Score 24; DB 1; Length 146;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 84 VSGSV 88

RESULT 14

O9RYV8 PRELIMINARY; PRT; 149 AA.

AC O9RYV8: 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DE 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
GN CATION TRANSPORT SYSTEM PROTEIN, PUTATIVE.  
GN DR0882.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Delnococcus group; Delnococcales; Deinococcus.  
CC NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-R1.  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Hatt D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Yamathevan J.J., Lam P., McDonald L., Uterback T., Zaleski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1.";  
RL Science 286:1571-1577(1999).  
DR EMBL; AE001941; AAF10455.1; -  
DR TIGR; DR0882; -  
DR InterPro: IPR002757; -  
DR Pfam: PF01898; DUF67; 1.  
SQ SEQUENCE 149 AA: 15296 MW: 3E107E0E63C86CF6 CRC64;

Query Match Best Local Similarity 100.0%; Score 24; DB 2; Length 149;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 13 VSGSV 17

RESULT 15

O9Y4T1 PRELIMINARY; PRT; 150 AA.

AC O9Y4T1: 01-NOV-1999 (TREMblrel. 12, Created)  
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
DE 01-NOV-1999 (TREMblrel. 12, Last annotation update)  
GN HYPOTHELICAL 16.3 KDA PROTEIN (FRAGMENT).  
GN DKF2P564N1362.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
CC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RA Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL080092; CAB45707.1; -  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 150 AA: 16324 MW: 8185875EC7D724E8 CRC64;

Query Match Best Local Similarity 100.0%; Score 24; DB 4; Length 150;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 144 VSGSV 148

Search completed: September 6, 2001, 16:49:53  
Job time: 737 sec

Fri Sep 7 10:57:57 2001

us-09-603-713-24.rspt

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:51:08 ; Search time 72.75 Seconds  
(without alignments)

2.354 Million cell updates/sec

Title: US-09-603-713-24

Perfect score: 24

Sequence: 1 VSGSV 5

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues  
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	105	THIO_TREPA	083889 treponema p
2	24	100.0	109	RLAI_ARTSA	P02402 artemia sal
3	24	100.0	218	SNL3_HUMAN	043761 homo sapien
4	24	100.0	221	R51L_ARCFU	028184 archaeoglob
5	24	100.0	235	H47_STEPL	041348 stellaria l
6	24	100.0	255	AC24_STRPO	P46106 streptomyce
7	24	100.0	259	YC09_PYRHO	058774 pyrococcus
8	24	100.0	265	PROC_AQUAE	066553 aquifex aeo
9	24	100.0	309	ERI_HEVBR	Q39563 hevea brasl
10	24	100.0	310	TATC_MYCLE	P54078 mycobacteri
11	24	100.0	319	K6PF_BACST	P00512 bacillus st
12	24	100.0	319	K6PF_BACSU	034529 bacillus su
13	24	100.0	327	QOR_ECOLI	P28304 escherichia
14	24	100.0	346	GARD_ECOLI	P37190 escherichia
15	24	100.0	354	DDL_SYNY3	P73632 synechocyst
16	24	100.0	373	DDL_MYCSM	Q92900 mycobacteri
17	24	100.0	391	EFPU_BACSU	P04220 homo sapien
18	24	100.0	396	EFPU_BACSU	P33166 bacillus su
19	24	100.0	399	EFPU_CAMOTE	069303 campylobact
20	24	100.0	422	HEM1_PSEAE	P42807 pseudomonas
21	24	100.0	426	HEM1_CHLPN	Q927M4 chlamydia p
22	24	100.0	433	DHOM_BACSU	P19582 bacillus su
23	24	100.0	450	PX8A_HUMAN	Q06710 homo sapien
24	24	100.0	454	MUC_HUMAN	P01871 homo sapien
25	24	100.0	469	VL2_BPY1	P03109 bovine papl
26	24	100.0	470	FUCK_HAEIN	P44399 haemophilus
27	24	100.0	481	6PGD_CERCA	P41570 ceratilis c
28	24	100.0	481	6PGD_DROME	P41572 drosophila
29	24	100.0	481	6PGD_DROSI	P41573 drosophila
30	24	100.0	482	6PGD_HUMAN	P52209 homo sapien
31	24	100.0	482	6PGD_SHEEP	P00349 ovls aries
32	24	100.0	485	6PGD_CUNEL	O60037 Cunningham
33	24	100.0	489	6PG1_YEAST	P38720 saccharomyc

34	24	100.0	492	1	6PG2_YEAST	P53319 saccharomyc
35	24	100.0	492	1	6PGD_SCHPO	P78612 schizosach
36	24	100.0	503	1	VP57_BPY	P52638 borra disca
37	24	100.0	517	1	6PGD_CANAL	013387 candida alb
38	24	100.0	568	1	HEMA_SY41	P25180 simian vtru
39	24	100.0	572	1	ERGI_MOUSE	P52019 mus musculi
40	24	100.0	607	1	YA27_SCHPO	Q09698 schizosach
41	24	100.0	667	1	EM70_YEAST	P32802 saccharomyc
42	24	100.0	673	1	Y552_HUMAN	O60289 homo sapien
43	24	100.0	703	1	PALY_BROFI	042609 bromheadia
44	24	100.0	710	1	PAL3_PRAVU	P19143 phaseolus v
45	24	100.0	744	1	NUAM_NEUCR	P24918 neurospora

## ALIGNMENTS

```

RESULT 1
THIO_TREPA
ID THIO_TREPA STANDARD: PRT; 105 AA.
AC 083889:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE THIOREDOXIN (TRX).
GN TRXA OR TP0919.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxId=160;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS.
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.R., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uitterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.,
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
CC -!- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS
CC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001260; AAC65875.1; -.
CC TIGR: TP0919; -.
DR InterPro: IPR000063; -.
DR Pfam: PF00085; Thioredox.
DR PRINTS: PR00421; THIOREDOXIN.
DR PROSITE: PS00194; THIOREDOXIN; 1.
KW Redox-active center; Electron transport.
FT DISULFID 30
FT SEQUENCE 105 AA; 11391 MM; 637AB85DAFE867 CRC64;
SQ

```

Query Match 100.0%; Score 24; DB 1; Length 105;

Best local Similarity 100.0%; Pred. No. 97;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
|||||

Db 47 VGSV 51

RESULT 2

RL1\_ARTSA STANDARD; PRT; 109 AA.

AC P02402;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-APR-1988 (Rel. 07, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE 60S ACIDIC RIBOSOMAL PROTEIN P1 (EL12'/EL12'-P).

OS Artemia salina (Brine shrimp).

OC Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda; Anostraca; Artemiidae; Artemia.

NCBI\_TaxID=85549;

OX [1]

RP SEQUENCE OF 1-106 FROM N.A.

RX MEDLINE=85230659; PubMed=3839187;

RA Maassen J.A., Schop E.N., Brands J.H.G.M., van Hemert F.J., Lensstra J.A., Moeller W.;

RT "Molecular cloning and analysis of cDNA sequences for two ribosomal proteins from Artemia. The coordinate expression of genes for ribosomal proteins and elongation factor 1 during embryogenesis of Artemia.";

RT Eur. J. Biochem. 149:609-616(1985).

RL [2]

RP SEQUENCE.

RA Amos R., Pluijms W.J.M., Kriek J., Moeller W.;

RT "The primary structure of protein el12'/el12'-P from the large subunit of Artemia salina ribosomes.";

RL FEBS Lett. 146:143-147(1982).

CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF PROTEIN SYNTHESIS.

CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL SUBUNIT.

CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X02633; CAA26480.1; -

DR PIR: B25208; R6SSP2.

DR InterPro: IPR001813; -

DR Pfam: PF00428; 60s\_ribosomal; 1.

KW Ribosomal protein; phosphorylation.

KW INIT\_MET 0 0

FT MOD\_RES 1 1 BLOCKED:

FT MOD\_RES 96 96 PHOSPHORYLATION (PARTIAL, IN EL12'P).

FT CONFLICT 8 8 MISSING (IN REF. 2).

FT CONFLICT 21 21 D -> V (IN REF. 2).

FT CONFLICT 42 42 W -> S (IN REF. 2).

FT CONFLICT 109 AA; 11407 MW; DE21B2AEBACACB7E CRC64;

SEQUENCE

Query Match 100.0%; Score 24; DB 1; Length 109;

Best Local Similarity 100.0%; Pred. No. 1e+02; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0;

OY 1 VGSV 5

Db 62 VGSV 66

RESULT 3

SN3\_HUMAN STANDARD; PRT; 218 AA.

ID SN3\_HUMAN

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE SYNAPTOGYRIN 3.

GN SYN3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OC NCBI\_TaxID=9606;

OX [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98430994; PubMed=9760194;

RA Kedra D., Pan H.-Q., Seroussi E., Fransson I., Guillaud C., Collins J.E., Dunham I., Blennow E., Roe B.A., Plehl F., Dumanski J.P.;

RT "Characterization of the human synaptogyrin gene family.";

RL Hum. Genet. 103:131-141(1998).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN AND PLACENTA.

CC -1- SIMILARITY: BELONGS TO THE SYNAPTOGYRIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AJ002309; CAA05326.1; -

DR MIM: 603927; -

KW Transmembrane.

FT TRANSMEM 19 39 POTENTIAL.

FT TRANSMEM 59 79 POTENTIAL.

FT TRANSMEM 94 114 POTENTIAL.

FT TRANSMEM 137 157 POTENTIAL.

FT SEQUENCE 218 AA; 23517 MW; 436574849EBDC06 CRC64;

Query Match 100.0%; Score 24; DB 1; Length 218;

Best Local Similarity 100.0%; Pred. No. 2e+02; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0;

OY 1 VGSV 5

Db 187 VGSV 191

RESULT 4

R51L\_ARCFU STANDARD; PRT; 221 AA.

ID R51L\_ARCFU

AC 028184;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DE PROBABLE DNA REPAIR PROTEIN RAD51-LIKE.

GN AF2096.

OS Archaeoglobus fulgidus.

OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae; Archaeoglobus.

OC NCBI\_TaxID=2234;

OX [1]

RP SEQUENCE FROM N.A.

RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;

RX MEDLINE=98049343; PubMed=9389475;

RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D., Richardson D.L., Kerlavage A.R., Graham D.E., Kyrleides N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uitterback T., Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,

RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Moase C.R.,  
 RA Venter J.C.:  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon *Archaeoglobus fulgidus*.";  
 CC Nature 390:364-370(1997).  
 CC -1- FUNCTION: PROBABLY INVOLVED IN DNA REPAIR.  
 CC -1- SIMILARITY: BELONGS TO THE RECA-LIKE PROTEINS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE000959; AAB89159.1; -.  
 CC TIGR: AF2096; -.  
 CC InterPro: IPR001553; -.  
 CC PRINTS: PRO0142; RECA.  
 CC PROSITE: PSS0162; RECA\_2; 1.  
 CC DNA damage; DNA repair; ATP-binding.  
 CC NP\_BIND 31 38 ATP (POTENTIAL).  
 CC FT SEQUENCE 221 AA: 24936 MW: 873363A3F7D1966F CRC64;  
 SQ  
 Query Match 100.0%; Score 24; DB 1; Length 221;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSV 5  
 Db 164 VSGSV 168  
 RESULT 5  
 H47\_STELP STANDARD: PRT; 235 AA.  
 ID H47\_STELP  
 AC 041348;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HYPOTHETICAL PROTEIN H47 (FRAGMENT).  
 OS Stellaria longipes.  
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 CC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;  
 CC Caryophyllales; Caryophyllaceae; Stellaria.  
 CC NCBI\_TaxID=19744;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC RP TISSUE-Leaf;  
 CC RA Zhang X.H.;  
 CC RL Submitted (MAR-1993) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0019 (SN2) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X71601; CAA50602.1; ALT\_INIT.  
 CC InterPro: IPR001852; -.  
 CC Pfam: PF01680; UPF0019; 1.  
 CC PROSITE: PSS01235; UPF0019; 1.  
 CC Hypothetical protein.  
 CC FT NON\_TER 1 1  
 CC SEQUENCE 235 AA: 25450 MW: 70095A53E2D0E0C CRC64;  
 SQ  
 Query Match 100.0%; Score 24; DB 1; Length 235;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSV 5  
 Db 164 VSGSV 168

Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSV 5  
 Db 173 VSGSV 177  
 RESULT 6  
 AC24\_STRCO STANDARD: PRT; 255 AA.  
 ID AC24\_STRCO  
 AC P46106;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PROBABLE ACTINORHODIN OPERON ACTIVATORY PROTEIN.  
 CC ACT11-4.  
 CC GN Streptomyces coelicolor.  
 CC OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC OC Actinomycetales; Streptomycineae; Streptomyetaceae; Streptomyces.  
 CC NCBI\_TaxID=1902;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC RP MEDLINE-91347376; PubMed-1878971;  
 CC RA Fernandez-Moreno M.A., Caballero J.L., Hopwood D.A., Malpartida F.;  
 CC "The act cluster contains regulatory and antibiotic export genes;  
 CC direct targets for translational control by the bldA tRNA gene of  
 CC Streptomyces.";  
 CC RL Cell 66:769-780(1991).  
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF ACTINORHODIN BIOSYNTHETIC  
 CC GENES.  
 CC -1- SIMILARITY: BELONGS TO THE AFSR/DNRI/REDD FAMILY OF REGULATORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M64683; AAA26692.1; -.  
 CC DR Antibiotic biosynthesis; Transcription regulation; Activator.  
 CC KW SEQUENCE 255 AA: 28762 MW: 59E1E749C37F3AAD CRC64;  
 SQ  
 Query Match 100.0%; Score 24; DB 1; Length 255;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSV 5  
 Db 41 VSGSV 45  
 RESULT 7  
 YC09\_PYRHO STANDARD: PRT; 259 AA.  
 ID YC09\_PYRHO  
 AC 058974;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HYPOTHETICAL PROTEIN PH1209.  
 CC GN PH1209 OR PHBK045.  
 CC OS Pyrococcus horikoshii.  
 CC OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 CC NCBI\_TaxID=53953;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC RP STRAIN-OT3;  
 CC RC MEDLINE-98344137; PubMed-9679194;  
 CC RX Kawarabayashi Y., Sawada M., Horikawa H., Hatake Y., Hino Y.,  
 CC Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,

RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Masuchi Y., Shizuya H., Kikuchi H.,  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii 033.";  
 RL DNA Res. 5:55-76(1998).  
 CC -1- SIMILARITY: BELONGS TO THE BTPA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AP000005; BAA30309.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 259 AA; 28209 MW; 6E95ADPD584D1913 CRC64;

Query Match 100.0%; Score 24; DB 1; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGSGV 5  
 |||||  
 Db 206 VGSGV 210

RESULT 8  
 PROC\_AQUAE STANDARD; PRT; 265 AA.  
 ID PROC\_AQUAE STANDARD; PRT; 265 AA.  
 AC 066553;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PYROLINE-5-CARBOXYLATE REDUCTASE (EC 1.5.1.2) (P5CR) (P5C REDUCTASE).  
 GN PROC OR AO\_106.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
 CX NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5;  
 RX MEDLINE=98196666; PubMed=9537320;  
 RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus.";  
 RL Nature 392:353-358(1998).  
 CC -1- CATALYTIC ACTIVITY: L-PYROLINE + NAD(P)(+) -> L-PYROLINE-5-  
 CC CARBOXYLATE + NAD(P)H.  
 CC -1- PATHWAY: TERMINAL (THIRD) STEP IN PYROLINE BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO THE PYROLINE-5-CARBOXYLATE REDUCTASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE000676; AAC0504.1; -  
 DR InterPro: IPR000304; -  
 DR Pfam: PF01089; P5CR; 1.  
 DR PROSITE: PS00521; P5CR; 1.  
 KW Oxidoreductase; Proline biosynthesis; NADP.  
 SQ SEQUENCE 265 AA; 28814 MW; 32DD6B49EDDCFA03 CRC64;

Query Match 100.0%; Score 24; DB 1; Length 265;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGSGV 5  
 |||||  
 Db 125 VGSGV 129

RESULT 9  
 ERL HEVER STANDARD; PRT; 309 AA.  
 ID ERL HEVER STANDARD; PRT; 309 AA.  
 AC Q39963;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE ETHYLENE-INDUCIBLE PROTEIN HEVER.  
 DE ERI.  
 GN Hevea brasiliensis (Para rubber tree).  
 OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
 OC Malpighiales; Euphorbiaceae; Hevea.  
 CX NCBI\_TaxID=3981;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Latexifier;  
 RX MEDLINE=96017624; PubMed=7579163;  
 RA Sivasubramanian S., Vanniasingham V.M., Tan C.T., Chua N.H.;  
 RA "Characterisation of HEVER, a novel stress-induced gene from Hevea  
 brasiliensis.";  
 RL Plant Mol. Biol. 29:173-178(1995).  
 CC -1- INDUCTION: BY STRESS TREATMENT WITH SALICYLIC ACID AND ETHEPHON.  
 CC -1- SIMILARITY: BELONGS TO THE UPF0019 (SNZ) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M88254; AAA91063.1; -  
 DR InterPro: IPR001852; -  
 DR Pfam: PF01680; UPF0019; 1.  
 DR PROSITE: PS01235; UPF0019; 1.  
 SQ SEQUENCE 309 AA; 33133 MW; AA3867583905403 CRC64;

Query Match 100.0%; Score 24; DB 1; Length 309;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGSGV 5  
 |||||  
 Db 250 VGSGV 254

RESULT 10  
 TATC MYCLE STANDARD; PRT; 310 AA.  
 ID TATC MYCLE STANDARD; PRT; 310 AA.  
 AC PS4078; Q95373;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE SEC-INDEPENDENT PROTEIN TRANSLOCASE PROTEIN TATC HOMOLOG.  
 GN TATC OR MTPB OR MCB2533.28 OR U2126A OR B2126.C1.183.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 CX NCBI\_TaxID=1769;

```

RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
RT
[2]
RP SEQUENCE FROM N.A.
RA Hamlin N., Churcher C.M., James K.D., Parkhill J., Barrell B.G.,
RA Rajandream M.A.;
RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR CORRECT LOCALIZATION OF PRECURSOR PROTEINS
CC BEARING SIGNAL PEPTIDES WITH THE TWIN ARGININE CONSERVED MOTIF
CC S/T-R-R-X-F-L-K. THIS SEC-INDEPENDENT PATHWAY IS TERMED TAT FOR
CC TWIN-ARGININE TRANSLLOCATION SYSTEM. THIS SYSTEM MAINLY TRANSPORTS
CC PROTEINS WITH BOUND COFACTORS THAT REQUIRE FOLDING PRIOR EXPORT (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE TATC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U00017; AAA17191.1; ALT_INIT.
CC DR EMBL: AL035310; CAA22942.1; -.
CC DR InterPro: IPR002033; -.
CC DR Pfam: PF00902; UPF0032; 1.
CC DR PROSITE: PS01218; TATC; 1.
CC KM Transport; Protein transport; Translocation; Transmembrane.
CC FT TRANSMEM 38 58 POTENTIAL.
CC FT TRANSMEM 109 129 POTENTIAL.
CC FT TRANSMEM 142 163 POTENTIAL.
CC FT TRANSMEM 163 182 POTENTIAL.
CC FT TRANSMEM 190 210 POTENTIAL.
CC FT TRANSMEM 229 249 POTENTIAL.
CC FT TRANSMEM 250 270 POTENTIAL.
CC SQ SEQUENCE 310 AA; 34535 MW; 743625AD9C3E96F CRC64;

Query Match 100.0%; Score 24; DB 1; Length 310;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 0; Incls 0; Gaps 0;
OY 1 VSGV 5
DB 173 VSGV 177

RESULT 11
K6PF_BACST STANDARD; PRT; 319 AA.
ID K6PF_BACST
AC P00512;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE 6-PHOSPHOFRUCTOKINASE (EC 2.7.1.11) (PHOSPHOFRUCTOKINASE)
DE (PHOSPHOHEXOKINASE).
DE PFKA OR PFK.
GN BACST
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-87277434; PubMed-2956156;
RT French B.A., Chang S.H.;
RT "Nucleotide sequence of the phosphofructokinase gene from Bacillus
RT stearothermophilus and comparison with the homologous Escherichia
RT coli gene."
RL Gene 54:65-71(1987).

```

```

RN [2]
RP SEQUENCE.
RX MEDLINE-81003881; PubMed-6447595;
RA Kolb E., Hudson P.J., Harris J.I.;
RT "Phosphofructokinase: complete amino-acid sequence of the enzyme from
RT Bacillus stearothermophilus."
RL Eur. J. Biochem. 108:587-597(1980).
RN [3]
RP SEQUENCE OF 300-319 FROM N.A.
RX MEDLINE-93170322; PubMed-8436141;
RA Sakai H., Ohta T.;
RT "Molecular cloning and nucleotide sequence of the gene for pyruvate
RT kinase of Bacillus stearothermophilus and the production of the
RT enzyme in Escherichia coli. Evidence that the genes for
RT phosphofructokinase and pyruvate kinase constitute an operon."
RL Eur. J. Biochem. 211:851-859(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE-7919719; PubMed-156307;
RA Evans P.R., Hudson P.J.;
RT "Structure and control of phosphofructokinase from Bacillus
RT stearothermophilus."
RL Nature 279:500-504(1979).
CC -1- CATALYTIC ACTIVITY: ATP + D-FRUCTOSE 6-PHOSPHATE -> ADP +
CC D-FRUCTOSE 1,6-BISPHOSPHATE.
CC -1- ENZYME REGULATION: BY ADP AND PHOSPHOENOLPYRUVATE, IN CONTRAST TO
CC THE EUKARYOTIC ENZYME, WHICH IS REGULATED BY AMP, ATP, AND
CC CITRATE.
CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- MISCELLANEOUS: THE 6-P GROUP OF FRUCTOSE 6-PHOSPHATE (F6P) IS
CC BOUND BY HIS-249 FROM ONE CHAIN AND ARG-162 AND ARG-243 FROM AN
CC ADJACENT CHAIN.
CC -1- MISCELLANEOUS: THE BETA-PHOSPHATE GROUP OF ADP IS BOUND BY ARG-154
CC FROM ONE CHAIN AND ARG-21 AND ARG-25 FROM AN ADJACENT CHAIN.
CC -1- MISCELLANEOUS: THE PHOSPHATE GROUP OF PHOSPHOENOLPYRUVATE APPEARS
CC TO BIND IN THE SAME PLACE AS THE BETA-PHOSPHATE GROUP OF ADP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M15643; AAA22656.1; -.
CC DR EMBL: D13095; BAA02405.1; -.
CC DR PIR: A00605; KIBSFF.
CC DR PIR: A27474; A27474.
CC DR PDB: 3PER; 15-JUL-90.
CC DR PDB: 4PER; 15-JUL-90.
CC DR PDB: 6PER; 11-JUL-96.
CC DR InterPro: IPR000023; -.
CC DR Pfam: PF00365; PERK; 1.
CC DR PRINTS: PRO0476; PHFRCTKINASE.
CC DR PROSITE: PS00433; PHOSPHOFRUCTOKINASE; 1.
CC KW kinase; Transferase; Glycolysis; ATP-binding; 3D-structure.
CC FT DOMAIN 1 1 134
CC FT DOMAIN 136 253
CC FT DOMAIN 254 304
CC FT DOMAIN 305 319
CC FT BINDING 21 21 BETA-PHOSPHATE GROUP OF ADP.
CC FT BINDING 25 25 BETA-PHOSPHATE GROUP OF ADP.
CC FT METAL 103 103 MAGNESIUM (POTENTIAL).
CC FT ACT_SITE 127 127 H BOND TO THE 1-HYDROXYL GROUP OF F6P.
CC FT BINDING 127 127 SUGAR RING.
CC FT BINDING 154 154 BETA-PHOSPHATE GROUP OF ADP.
CC FT BINDING 158 158 RIBOSE OF ADP.
CC FT BINDING 162 162 6-P GROUP OF FRUCTOSE 6-PHOSPHATE (F6P).
CC FT BINDING 169 169 SUGAR RING.
CC FT BINDING 171 171 ATP.
CC FT BINDING 215 215 RIBOSE OF ADP.

```

```

FT BINDING 221 221 SUGAR RING.
FT BINDING 243 243 6-P GROUP OF FRUCTOSE 6-PHOSPHATE (F6P).
FT BINDING 249 249 6-P GROUP OF FRUCTOSE 6-PHOSPHATE (F6P).
FT BINDING 252 252 SUGAR RING.
FT CONFLICT 12 12 D -> N (IN REF. 2).
FT CONFLICT 35 37 MISSING (IN REF. 2).
FT CONFLICT 43 43 G -> V (IN REF. 2).
FT CONFLICT 82 82 Q -> E (IN REF. 2).
FT CONFLICT 95 95 E -> Q (IN REF. 2).
FT CONFLICT 225 225 G -> L (IN REF. 2).
FT STRAND 3 9
FT TURN 14 15
FT HELIX 16 29
FT STRAND 30 31
FT STRAND 33 37
FT HELIX 40 45
FT TURN 46 47
FT STRAND 49 52
FT HELIX 54 57
FT TURN 58 59
FT TURN 61 62
FT TURN 67 68
FT TURN 74 75
FT HELIX 79 92
FT TURN 93 93
FT STRAND 96 101
FT HELIX 103 114
FT TURN 115 116
FT STRAND 119 124
FT TURN 131 132
FT STRAND 137 137
FT TURN 138 138
FT HELIX 139 159
FT TURN 160 161
FT STRAND 163 168
FT HELIX 175 183
FT TURN 184 185
FT STRAND 188 190
FT TURN 192 193
FT HELIX 198 210
FT TURN 211 212
FT STRAND 216 221
FT TURN 222 224
FT HELIX 227 238
FT STRAND 242 246
FT HELIX 248 252
FT HELIX 258 276
FT TURN 277 278
FT STRAND 282 287
FT TURN 288 289
FT STRAND 290 295
FT HELIX 296 299
FT TURN 300 301
FT HELIX 308 317
FT TURN 318 318
SQ SEQUENCE 319 AA; 34119 MW; EE96BD39BA30712B CRC64;

```

Query Match 100.0%; Score 24; DB 1; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
 DB 224 VSGSV 228

RESULT 12  
 K6PF\_BACSU STANDARD; PRT; 319 AA.  
 AC 034528;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)

```

DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 6-PHOSPHOFRUCTOKINASE (EC 2.7.1.11) (PHOSPHOFRUCTOKINASE)
GN PEKA OR PFK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
ON NCBI_TaxId=1423;
RX SEQUENCE FROM N.A.
RP MEDLINE=98048467; PubMed=9387221;
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequencing and functional annotation of the Bacillus subtilis genes
in the 200 kb rnh-dnaB region."
RL Microbiology 143:3431-3441(1997).
CC -1- CATALYTIC ACTIVITY: ATP + D-FRUCTOSE 6-PHOSPHATE = ADP +
D-FRUCTOSE 1,6-BISPHOSPHATE.
CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF008220; AAC00342.1; -.
CC DR EMBL; Z99118; CAB14879.1; -.
CC DR HSSP; P00512; 6PFK.
CC DR Subtilist; BG12644; PFKA.
CC DR InterPro; IPR000023; -.
CC DR Pfam; PF00365; PFK; 1.
CC DR PRINTS; PR00476; PHERCTKINASE.
CC DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.
CC KW Kinase; Transferase; Glycolysis.
SQ SEQUENCE 319 AA; 34254 MW; 706CBCTF9BCFCDC CRC64;

```

Query Match 100.0%; Score 24; DB 1; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
 DB 224 VSGSV 228

RESULT 13  
 QOR\_ECOLI STANDARD; PRT; 327 AA.  
 AC P28304;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE QUINONE OXIDOREDUCTASE (EC 1.6.5.5) (NADPH:QUINONE REDUCTASE) (ZETA-  
 CRYSTALLIN HOMOLOG PROTEIN).  
 GN QOR OR HCV.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 ON NCBI\_TaxId=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA Dixon N.E., Lilley P.E.;  
 RL Submitted (SEP-1992) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94089397; PubMed=8265357;  
 RA Bialtner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,

```

RA Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes.";
RN Nucleic Acids Res. 21:5408-5417(1993).
RL [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE-95326140; PubMed-7602590;
RT Thorn J.M., Barton J.D., Dixon N.E., Ollis D.L., Edwards K.J.;
RT "Crystal structure of Escherichia coli QOR quinone oxidoreductase
RT complexed with NADPH.";
RL J. Mol. Biol. 249:785-799(1995).
CC -1- CATALYTIC ACTIVITY: NADH + QUINONE = NADP(+) + SEMIQUINONE.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY. QUINONE OXIDOREDUCTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: I02312; AAA3691.1; -
CC EMBL: U00006; AAC3145.1; -
CC EMBL: AE000478; AAC7021.1; -
CC PIR: S45529; S45529.
CC PDB: 1OOR: 03-JUN-95.
CC Ecogene: EGI1492; gor.
CC InterPro: IPR002085; -.
CC InterPro: IPR002364; -.
CC Pfam: PF00107; adh_zinc; 1.
CC PROSITE: PS01162; QOR_ZETA.CRYSTAL; 1.
CC Oxidoreductase: NADP: zinc: 3D-structure.
SQ SEQUENCE 327 AA: 35172 MW: A6355B12DBA513B2 CRC64;

Query Match 100.0%; Score 24; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5
Db 72 VSGSV 76

RESULT 14
GATD_ECOLI STANDARD; PRT; 346 AA.
ID GATD_ECOLI
AC P37190; P76410;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GALACTITOL-1-PHOSPHATE 5-DEHYDROGENASE (EC 1.1.1.251).
GN GATD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-EC3132; PubMed-7772602;
RX MEDLINE-95290497; PubMed-7772602;
RT Nobelmann B., Lengeler J.M.;
RT "Sequence of the gat operon for galactitol utilization from a
RT wild-type strain EC3132 of Escherichia coli.";
RL Biochim. Biophys. Acta 1262:69-72(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RX Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

```

```

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474(1997).
RL [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12; PubMed-9097040;
RX Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RX Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RX Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RX Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RX Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,
RX Takeda J., Takemoto K., Wada C., Yamamoto Y., Horuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -1- FUNCTION: CONVERTS GALACTITOL-1-PHOSPHATE TO TAGATOSE 6-PHOSPHATE.
CC -1- CATALYTIC ACTIVITY: GALACTITOL-1-PHOSPHATE + H2O -> L-TAGATOSE
CC 6-PHOSPHATE + NADH.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X79837; CAA56231.1; -
CC EMBL: AE000298; AAC75152.1; -
CC EMBL: D90846; BAA15943.1; -
CC EMBL: D90847; BAA15954.1; -
CC HSP: P07846; ISDG.
CC Ecogene: EGI2417; gald.
CC InterPro: IPR002085; -.
CC InterPro: IPR002328; -.
CC Pfam: PF00107; adh_zinc; 1.
CC PROSITE: PS00059; ADH_ZINC; 1.
DR GALACTITOL metabolism; Oxidoreductase; NAD; zinc.
KW METAL
FT METAL 38
FT METAL 59
FT METAL 89
FT METAL 92
FT METAL 95
FT METAL 103
FT METAL 144
FT METAL 213
FT METAL 222
FT METAL 226
FT METAL 227
FT METAL 271
FT METAL 327
FT METAL 329
SQ SEQUENCE 346 AA: 37390 MW: B35A2B004E5A4032 CRC64;

Query Match 100.0%; Score 24; DB 1; Length 346;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5
Db 68 VSGSV 72

RESULT 15
DDL_SYNY3 STANDARD; PRT; 354 AA.
ID DDL_SYNY3
AC P73632;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)

```

DE D-ALANINE--D-ALANINE LIGASE (EC 6.3.2.4) (D-ALANYLALANINE SYNTHETASE)  
 DE (D-ALA-D-ALA LIGASE).  
 GN DDL OR DDLA OR SLR1874.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_taxid=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hikosawa M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K.,  
 RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,  
 RA Yamada M., Yasuda M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions";  
 RL DNA Res. 3:109-136(1996).  
 CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + D-ALANINE + D-ALANINE = ADP +  
 CC ORTHOPHOSPHATE + D-ALANYL-D-ALANINE.  
 CC -1- PATHWAY: ALONG WITH ALANINE RACEMASE, IT MAKES UP THE D-ALANINE  
 CC BRANCH OF THE PEPTIDOGLYCAN BIOSYNTHETIC ROUTE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE D-ALANINE--D-ALANINE LIGASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D90908: BAA1677.1; .  
 DR HSRP: P07862: 2DLN.  
 DR InterPro: IPR000291.  
 DR Pfam: PF01820; Data\_Data\_ligas; 1.  
 DR PROSITE: PS00843; DATA\_DATA\_LIGASE\_1; 1.  
 DR PROSITE: PS00844; DATA\_DATA\_LIGASE\_2; 1.  
 KW Ligase; Cell wall; Peptidoglycan synthesis.  
 SO SEQUENCE 354 AA; 38774 MW; 871B82362A47315A CRC64;

Query Match 100.0%; Score 24; DB 1; Length 354;  
 Best Local Similarity 100.0%; Pred. No. 3; 1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSV 5  
 |||||  
 DB 115 VSGSV 119

Search completed: September 6, 2001, 16:51:09  
 Job time: 813 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:45:51 : Search time 134.15 seconds  
(without alignments)  
2.839 Million cell updates/sec

Title: US-09-603-713-24

Perfect score: 24

Sequence: 1 VSGSV 5

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	100.0	105	2	D71265
2	24	100.0	110	1	R6SSP2
3	24	100.0	113	2	C81525
4	24	100.0	113	2	E72013
5	24	100.0	113	2	F86610
6	24	100.0	122	2	S72797
7	24	100.0	146	2	F72505
8	24	100.0	147	2	I54245
9	24	100.0	149	2	H75464
10	24	100.0	150	2	T12479
11	24	100.0	165	2	E84346
12	24	100.0	180	2	T01045
13	24	100.0	190	2	S71492
14	24	100.0	196	2	C84919
15	24	100.0	200	2	S77807
16	24	100.0	218	2	S36974
17	24	100.0	221	1	H69511
18	24	100.0	222	2	S33204
19	24	100.0	227	2	D84058
20	24	100.0	251	2	T44757
21	24	100.0	255	2	D40046
22	24	100.0	259	2	C71064
23	24	100.0	265	2	F70315
24	24	100.0	267	2	G84401
25	24	100.0	283	2	A56721
26	24	100.0	283	2	T18662
27	24	100.0	287	2	G72475
28	24	100.0	306	2	T23337
29	24	100.0	309	2	S60047

## ALIGNMENTS

30	24	100.0	309	2	T01255	probable ethylene-
31	24	100.0	310	1	E64525	nodulation protein
32	24	100.0	310	2	E71981	probable sugar nuc
33	24	100.0	317	2	S72851	hypothetical prote
34	24	100.0	319	1	K185F	hypothetical prote
35	24	100.0	319	2	A69675	6-phosphofructokin
36	24	100.0	319	2	F72750	hypothetical prote
37	24	100.0	327	1	S45529	quinone oxidoreduc
38	24	100.0	327	2	E86098	hypothetical prote
39	24	100.0	328	2	B86314	hypothetical prote
40	24	100.0	328	2	T01373	hypothetical prote
41	24	100.0	329	2	T10203	hypothetical prote
42	24	100.0	332	2	F85058	hypothetical prote
43	24	100.0	339	2	D75529	D-alanine--D-alani
44	24	100.0	339	2	F72451	probable peptide t
45	24	100.0	342	2	A82968	alcohol dehydrogen

## RESULT 1

D71265  
probable thioresoxin (trx) - syphilis spirochete  
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999

C:Accession: D71265

R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M  
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A:Reference number: A71250; MUID:98332770

A:Accession: D71265  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-105 <COL>

A:Cross-references: GB:AE001260; GB:AE00520; NID:g3323232; PIDN:AAC65875.1; PID:g332

C:Genetics:

A:Gene: TP0919

C:Superfamily: thioresoxin; thioresoxin homology

Query Match 100.0%; Score 24; DB 2; Length 105;

Best Local Similarity 100.0%; Pred No. 1.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
Db 47 VSGSV 51

## RESULT 2

R6SSP2  
acidic ribosomal protein p1 - brine shrimp

N:Alternate names: ribosomal protein el12', ribosomal protein SL12ell

C:Species: Artemia salina (brine shrimp)

C:Date: 17-Dec-1982 #sequence\_revision 31-Mar-1992 #text\_change 24-Nov-1999

C:Accession: B25208; A02778

R:Maassen, J.A.; Schop, E.N.; Brands, J.H.G.M.; Van Hemert, F.J.; Molli

Eur. J. Biochem. 149, 609-616, 1985

A:Title: Molecular cloning and analysis of cDNA sequences for two ribosomal proteins

is of Artemia.

A:Reference number: A91146; MUID:85230659

A:Accession: B25208

A:Molecule type: mRNA

A:Residues: 1-107 <MA>

A:Cross-references: GB:X02633; NID:g5688; PIDN:CAA26480.1; PID:g5689

FEBS Lett. 146, 143-147, 1982

A:Title: The primary structure of protein el12'-p from the large subunit of Art

A:Reference number: A02778

A:Accession: A02778  
A:Molecule type: protein  
A:Residues: 2-8,10-21,'V',22-42,'S',44-110 <AM0>  
C:Superfamily: rat acidic ribosomal protein P1  
C:Keywords: blocked amino end, phosphoprotein; protein biosynthesis; ribosome  
F:2-110/Product: acidic ribosomal protein P1 #status experimental <MAT>  
F:2/Modified site: blocked amino end (Ala) (in mature form) #status experimental  
F:97/Binding site: phosphate (Ser) (covalent) (partial) #status experimental

Query Match 100.0%; Score 24; DB 1; Length 110;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 63 VSGSV 67

## RESULT 3

C81525

Conserved hypothetical protein CP0900 [imported] - Chlamydia pneumoniae (strain AR39)

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000

C:Accession: C81525

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Ginn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.

A:Reference number: AB1500; MUID:20150255

A:Accession: C81525

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-113 <REA>

A:Cross-references: GB:AE002249; GB:AE002161; NID:g7189811; PIDN:AAF38686.1; PID:g7189811

A:Experimental source: strain AR39, HL cells

C:Genetics:

A:Gene: CP0900

Query Match

Best Local Similarity 100.0%; Score 24; DB 2; Length 113;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 58 VSGSV 62

## RESULT 4

E72013

CT809 hypothetical protein - Chlamydia pneumoniae (strain CWL029)

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000

C:Accession: E72013

R:Kaltman, S.; Mitchell, W.; Marathe, R.; Lamme, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606

A:Accession: E72013

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-113 <ARN>

A:Cross-references: GB:AE001676; GB:AE001363; NID:g4377283; PIDN:AAD19096.1; PID:g437728

A:Experimental source: strain CWL029

C:Genetics:

A:Gene: CP0960

Query Match

Best Local Similarity 100.0%; Score 24; DB 2; Length 113;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 58 VSGSV 62

## RESULT 5

F86610

CT809 hypothetical protein [imported] - Chlamydia pneumoniae (strain J138)

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001

C:Accession: F86610

R:Shitai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.

A:Reference number: AB6491; MUID:20330349

A:Accession: F86610

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-113 <STO>

A:Cross-references: GB:BA000008; NID:g8979334; PIDN:BA09168.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

A:Gene: CPJ0960

Query Match

Best Local Similarity 100.0%; Score 24; DB 2; Length 113;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 58 VSGSV 62

## RESULT 6

S72797

probable K+ transport protein tp2 - Mycobacterium leprae

N:Alternate names: B1549, C2.206 protein

C:Species: Mycobacterium leprae

C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 23-Mar-2001

C:Accession: S72797

R:Smith, D.R.; Robison, K. submitted to the EMBL Data Library, November 1993

A:Description: Mycobacterium leprae cosmid B1549.

A:Reference number: S72582

A:Accession: S72797

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-122 <SMI>

A:Cross-references: EMBL:U00014; NID:g466903; PIDN:AA050887.1; PID:g466912

Query Match

Best Local Similarity 100.0%; Score 24; DB 2; Length 122;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 34 VSGSV 38

## RESULT 7

F72505

hypothetical protein APE2020 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999

C:Accession: F72505

R:Kawarayashi, Y.; Hino, Y.; Horikawa, R.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Tawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J. DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A:Reference number: A72450; MUID:99310339

A:Accession: F72505

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-146 <RAW>  
A:Cross-references: DDBJ:AP000063; NID:g5105654; PIDN:BAAB81030.1; PID:d1044816; PID:g510  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE2020

Query Match 100.0%; Score 24; DB 2; Length 146;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 84 VSGSV 88

RESULT 8  
I54245  
laminin M - rat (fragment)  
C:Species: Rattus sp. (rat)  
C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 04-Sep-1998  
C:Accession: I54245  
R:Maher, J.J.; Tzagarakis, C.  
Hepatology 19, 764-770, 1994  
A>Title: Partial cloning of the M subunit of laminin from adult rat lipocytes: expressio  
A:Reference number: I54245; MUID:94164663  
A:Accession: I54245  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-147 <RES>  
A:Cross-references: GB:S72407; NID:g632888  
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 100.0%; Score 24; DB 2; Length 147;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 65 VSGSV 69

RESULT 9  
H75464  
probable cation transport system protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: H75464  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M., Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uutterback, T.; Zaleski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: H75464  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-149 <WHI>  
A:Cross-references: GB:AE001941; GB:AE000513; NID:g6458589; PIDN:AAFI0455.1; PID:g645859  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0882  
A:Map position: 1

Query Match 100.0%; Score 24; DB 2; Length 149;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5

DB 13 VSGSV 17  
|||||

RESULT 10  
T12479  
hypothetical protein DKFZP564N1362.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 23-Jul-1999  
C:Accession: T12479  
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
Submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z17526  
A:Accession: T12479  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-150 <BLU>  
A:Cross-references: EMBL:AL080092  
A:Experimental source: fetal brain; clone DKFZP564N1362  
C:Genetics:  
A:Note: DKFZP564N1362.1

Query Match 100.0%; Score 24; DB 2; Length 150;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 144 VSGSV 148

RESULT 11  
E84346  
hypothetical protein Vng1956h [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: E84346  
R:Ng, W.V.; Kennedy, S.P.; Mahatras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laaky  
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;  
A>Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: E84346  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-165 <STO>  
A:Cross-references: GB:AE004437; NID:g10581394; PIDN:AG20137.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG1956H

Query Match 100.0%; Score 24; DB 2; Length 165;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VSGSV 5  
|||||  
DB 52 VSGSV 56

RESULT 12  
T01045  
hypothetical protein YUP8H12R.28 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 22-Oct-1999  
C:Accession: T01045  
R:Rheologis, A.; Vysotskaya, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; K  
Oefner, P.; Davis, R.W.  
Submitted to the EMBL Data Library, May 1998  
A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.

A:Reference number: Z14227  
A:Accession: T01045  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-180 <THE>  
A:Cross-references: EMBL:AC002986; NID:q2494106; PID:g3152586; GSPDB:GN00059; ATSP:YUP8H  
C:Genetics:  
A:Gene: ATSP:YUP8H12R.28  
A:Map position: 1  
A:Introns: 55/1; 79/3; 111/2; 131/2; 141/3

Query Match 100.0%; Score 24; DB 2; Length 180;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGS GV 5  
DB 21 VGS GV 25

## RESULT 13

S71492  
ethylene-responsive protein 2 - Para rubber tree (fragment)  
C:Species: Hevea brasiliensis (Para rubber tree)  
C>Date: 09-Dec-1997 #sequence\_revision 06-Feb-1998 #text\_change 17-Nov-2000  
C:Accession: S71492  
R:Silvassubramaniam, S.; Vanniasingham, V.M.; Tan, C.T.; Chua, N.H.  
Plant Mol. Biol. 29, 173-178, 1995  
A:Title: Characterisation of HEVEA, a novel stress-induced gene from Hevea brasiliensis.  
A:Reference number: S60047; MIMD:96017624  
A:Accession: S71492  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-190 <STV>  
A:Cross-references: EMBL:M88255  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1996  
C:Superfamily: hypothetical protein H11647

Query Match 100.0%; Score 24; DB 2; Length 190;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGS GV 5  
DB 131 VGS GV 135

## RESULT 14

C84919  
hypothetical protein At2g47770 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: C84919  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MIMD:20083487  
A:Accession: C84919  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-196 <STO>  
A:Cross-references: GB:AE002093; NID:g3738290; PIDN:AAC63632.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g47770  
A:Map position: 2

Query Match 100.0%; Score 24; DB 2; Length 196;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGS GV 5  
DB 141 VGS GV 145

## RESULT 15

S77807  
Probable DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) chain A - Mycoplasma capri  
N:Alternate names: DNA gyrase; protein MC040  
C:Species: Mycoplasma capricolum  
C>Date: 09-Oct-1997 #sequence\_revision 31-Oct-1997 #text\_change 07-Dec-1999  
C:Accession: S77807; S46924  
R:Bork, P.; Ouzounis, C.; Casari, G.; Schneider, R.; Sander, C.; Dolan, M.; Gilbert,  
Mol. Microbiol. 16, 955-967, 1995  
A:Title: Exploring the Mycoplasma capricolum genome: a minimal cell reveals its phys  
A:Reference number: S77739; MIMD:96059641  
A:Accession: S77807  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-200 <BOR>  
A:Cross-references: EMBL:Z33035; NID:9514444; PIDN:CAA83715.1; PID:g530409  
A:Experimental source: ATCC 27343  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994  
C:Genetics:  
A:Genetic code: SGC3  
C:Keywords: Isomerase

Query Match 100.0%; Score 24; DB 2; Length 200;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGS GV 5  
DB 64 VGS GV 68

Search completed: September 6, 2001, 16:45:52  
Job time: 496 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2001, 16:39:37 : Search time 113.12 Seconds  
(without alignments)  
0.910 Million cell updates/sec

Title: US-09-603-713-24

Perfect score: 24

Sequence: 1 VSGGV 5

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents, AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/FCIUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	95	1 US-08-686-878A-30	Sequence 30, Appl
2	24	100.0	353	3 US-08-803-603-3	Sequence 3, Appl1
3	24	100.0	433	1 US-08-700-359-20	Sequence 20, Appl
4	24	100.0	503	3 US-08-803-603-1	Sequence 1, Appl1
5	24	100.0	503	3 US-08-369-822C-8	Sequence 8, Appl1
6	24	100.0	503	3 US-08-779-764A-26	Sequence 26, Appl
7	24	100.0	503	3 US-08-779-764A-27	Sequence 27, Appl
8	24	100.0	503	3 US-08-582-776C-8	Sequence 8, Appl1
9	24	100.0	503	3 US-08-434-831B-8	Sequence 8, Appl1
10	24	100.0	504	1 US-07-932-915-2	Sequence 2, Appl1
11	24	100.0	504	5 PCR-US91-05826-2	Sequence 2, Appl1
12	24	100.0	572	4 US-08-745-934-3	Sequence 3, Appl1
13	24	100.0	572	4 US-09-147-009-6	Sequence 6, Appl1
14	24	100.0	652	1 US-08-050-684-2	Sequence 2, Appl1
15	24	100.0	652	1 US-08-582-719-2	Sequence 2, Appl1
16	24	100.0	667	4 US-08-959-004-11	Sequence 11, Appl1
17	24	100.0	680	1 US-07-674-287B-2	Sequence 2, Appl1
18	24	100.0	680	2 US-08-436-900A-2	Sequence 4, Appl1
19	24	100.0	682	2 US-08-436-900A-4	Sequence 4, Appl1
20	24	100.0	778	6 5198347-4	Patent No. 5198347
21	24	100.0	1130	2 US-08-460-309-2	Sequence 2, Appl1
22	24	100.0	1130	2 US-08-125-077-2	Sequence 2, Appl1
23	24	100.0	1130	6 5444158-2	Patent No. 5444158
24	24	100.0	1996	2 US-08-804-227C-9	Sequence 9, Appl1
25	24	100.0	1996	2 US-08-804-198-3	Sequence 3, Appl1
26	24	100.0	3111	2 US-08-460-309-4	Sequence 4, Appl1
27	24	100.0	3111	2 US-08-125-077-4	Sequence 4, Appl1

28	24	100.0	15281	2 US-08-471-119A-2	Sequence 2, Appl1
29	23	95.8	40	2 US-08-576-626A-38	Sequence 38, Appl
30	23	95.8	100	3 US-09-039-859-6	Sequence 6, Appl1
31	23	95.8	102	3 US-09-039-859-4	Sequence 4, Appl1
32	23	95.8	143	3 US-09-039-859-7	Sequence 7, Appl1
33	23	95.8	307	3 US-09-039-859-11	Sequence 11, Appl
34	23	95.8	343	3 US-09-039-859-2	Sequence 2, Appl1
35	23	95.8	371	2 US-08-829-026A-6	Sequence 6, Appl1
36	23	95.8	401	2 US-08-591-079-6	Sequence 6, Appl1
37	23	95.8	401	2 US-08-591-079-6	Sequence 6, Appl1
38	23	95.8	485	1 US-08-362-512A-2	Sequence 2, Appl1
39	23	95.8	485	1 US-08-364-939-2	Sequence 2, Appl1
40	23	95.8	493	1 US-08-362-512A-4	Sequence 4, Appl1
41	23	95.8	493	4 US-08-964-939-4	Sequence 4, Appl1
42	23	95.8	496	4 US-08-949-637-2	Sequence 2, Appl1
43	23	95.8	496	4 US-09-291-488-2	Sequence 2, Appl1
44	23	95.8	560	2 US-08-928-692-53	Sequence 53, Appl
45	23	95.8	573	2 US-08-745-934-4	Sequence 4, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-686-878A-30  
: Sequence 30, Application US/0866878A  
: Patent No. 5708157  
: GENERAL INFORMATION:  
: APPLICANT: Jacobs, Kenneth  
: APPLICANT: McCoy, John  
: APPLICANT: Lavallee, Edward  
: APPLICANT: Racie, Lisa  
: APPLICANT: Merberg, David  
: APPLICANT: Treacy, Maurice  
: APPLICANT: Evans, Cheryl  
: APPLICANT: Spaulding, Vikki  
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
: NUMBER OF SEQUENCES: 71  
: CORRESPONDENCE ADDRESS:  
: ADDRESSER: Genetics Institute, Inc.  
: STREET: 87 CambridgePark Drive  
: CITY: Cambridge  
: STATE: Massachusetts  
: COUNTRY: U.S.A.  
: ZIP: 02140  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: OPERATING SYSTEM: IBM PC compatible  
: SOFTWARE: Patent Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/686,878A  
: FILING DATE:  
: CLASSIFICATION: 536  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Brown, Scott A.  
: REGISTRATION NUMBER: 32,724  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (617) 498-8224  
: TELEFAX: (617) 876-5851  
: INFORMATION FOR SEQ ID NO: 30:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 95 amino acids  
: TYPE: amino acid  
: STRANDEDNESS:  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-686-878A-30  
Query Match 100.0%; Score 24; DB 1; Length 95;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
Db 73 VSGSV 77

## RESULT 2

US-08-803-603-3  
; Sequence 3, Application US/08803603  
; Patent No. 6010700  
; GENERAL INFORMATION:  
; APPLICANT: Richt, Jurgen A.  
; TITLE OF INVENTION: Polypeptides Corresponding To The Amino Acid  
; TITLE OF INVENTION: Sequences of Proteins P57 or P9.5 of Borna Disease Virus, Nucle  
; TITLE OF INVENTION: Acid Fragments Coding Therefore and Their Use for Diagnostic  
; TITLE OF INVENTION: Immunization Purposes  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Felte & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/803,603  
; FILING DATE: 21-February-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EPA 96102575.6  
; FILING DATE: 21-February-1996

## ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 6010700man D.

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LEDER 208

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ. ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 353 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Peptide

FRAGMENT TYPE: Internal

US-08-803-603-3

Query Match 100.0%; Score 24; DB 3; Length 353;  
Best Local Similarity 100.0%; Pred. No. 1.3e+03;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
Db 226 VSGSV 230

## RESULT 3

US-08-700-359-20

; Sequence 20, Application US/08700359

; Patent No. 5766925

; GENERAL INFORMATION:

; APPLICANT: SUGIMOTO, MASAKAZU

; APPLICANT: USUDA, YOSHIHIRO

; APPLICANT: SUZUKI, TOMOKO

; APPLICANT: TANAKA, AKIKO

APPLICANT: MATSUI, HIROSHI

TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESSES:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS-DOS TEXT EDITOR

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/700,359

FILING DATE: 08-OCT-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP-35019

FILING DATE: 04-MAR-1994

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 10-819-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

INFORMATION FOR SEQ. ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 433 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-700-359-20

Query Match 100.0%; Score 24; DB 1; Length 433;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
Db 14 VSGSV 18

## RESULT 4

US-08-803-603-1

; Sequence 1, Application US/08803603

; Patent No. 6010700

; GENERAL INFORMATION:

; APPLICANT: Richt, Jurgen A.

; TITLE OF INVENTION: Polypeptides Corresponding To The Amino Acid

; TITLE OF INVENTION: Sequences of Proteins P57 or P9.5 of Borna Disease Virus, N

; TITLE OF INVENTION: Acid Fragments Coding Therefore and Their Use for Diagnosti

; TITLE OF INVENTION: Immunization Purposes

; NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Felte & Lynch

STREET: 805 Third Avenue

CITY: New York City

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/803,603

FILING DATE: 21-February-1997

CLASSIFICATION: 435



PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EPA 96102575.6  
FILING DATE: 21-February-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, NO. 6010700man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LEDER 208  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
FRAGMENT TYPE: Internal  
US-08-803-603-1

Query Match 100.0%; Score 24; DB 3; Length 503;  
Best Local Similarity 100.0%; Pred. NO. 1.7e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
11111  
DB 376 VSGSV 380

RESULT 5  
US-08-369-822C-8  
Sequence 8, Application US/08369822C  
Patent No. 6015660  
GENERAL INFORMATION:  
APPLICANT: Lipkin, W. I.  
APPLICANT: Brlese, Thomas  
APPLICANT: Kliche, Stefanie  
APPLICANT: Schneider, Patrick A.  
APPLICANT: Stiltz, Lothar  
APPLICANT: Schneemann, Anette  
TITLE OF INVENTION: Botna Disease Viral Sequences,  
TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
TITLE OF INVENTION: System Diseases  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski, L.L.P.  
STREET: 865 South Figueroa Street, 29th Floor  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90017-2571  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0  
SOFTWARE: ASCII (DOS) TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/369,822C  
FILING DATE: 06-JAN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Churchill, Margaret A. (Ph.D.)  
REGISTRATION NUMBER: 39,944  
REFERENCE/DOCKET NUMBER: 1279-194XX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 213/892-9200  
TELEFAX: 213/680-4518  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: unknown  
MOLECULE TYPE: Protein  
US-08-369-822C-8

Query Match 100.0%; Score 24; DB 3; Length 503;  
Best Local Similarity 100.0%; Pred. NO. 1.7e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
11111  
DB 376 VSGSV 380

RESULT 6  
US-08-779-764A-26  
Sequence 26, Application US/08779764A  
Patent No. 6057094  
GENERAL INFORMATION:  
APPLICANT: de la Torre, Juan C.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING  
TITLE OF INVENTION: OF HUMAN BORNA DISEASE VIRUS  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE  
STREET: 10550 No. 6057094th Torrey Pines Road, TPC-8  
CITY: La Jolla  
STATE: California  
COUNTRY: United States  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/779,764A  
FILING DATE: 16-DEC-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: TSRI 465.0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 784-2937  
TELEFAX: (619) 784-9399  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-779-764A-26

Query Match 100.0%; Score 24; DB 3; Length 503;  
Best Local Similarity 100.0%; Pred. NO. 1.7e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
11111  
DB 376 VSGSV 380

RESULT 7  
US-08-779-764A-27  
Sequence 27, Application US/08779764A  
Patent No. 6057094  
GENERAL INFORMATION:  
APPLICANT: de la Torre, Juan C.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING  
TITLE OF INVENTION: OF HUMAN BORNA DISEASE VIRUS  
NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:  
ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE  
STREET: 10550 No. 6057094th Torrey Pines Road, TPC-8  
CITY: La Jolla  
STATE: California  
COUNTRY: United States  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/779,764A  
FILING DATE: 16-DEC-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: TSRI 465.0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 784-9397  
TELEFAX: (619) 784-9399  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-779-764A-27

Query Match 100.0%; Score 24; DB 3; Length 503;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
DB 376 VSGSV 380

RESULT 8  
US-08-582-776C-8  
Sequence 8, Application US/08582776C  
Patent No. 6077510  
GENERAL INFORMATION:  
APPLICANT: Lipkin, W. I.  
APPLICANT: Briese, Thomas  
APPLICANT: Kliche, Stefanie  
APPLICANT: Schneider, Patrick A.  
APPLICANT: Stitz, Lothar  
APPLICANT: Schneemann, Anette  
TITLE OF INVENTION: Borina Disease Viral Sequences,  
TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
TITLE OF INVENTION: System Diseases  
NUMBER OF SEQUENCES: 61  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski, L.L.P.  
STREET: 865 South Figueroa Street, 29th Floor  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90017-2576  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: WINDOWS NT  
SOFTWARE: ASCII DOS TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/582,776C  
FILING DATE: 04-JAN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/369,822  
FILING DATE: 06-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/434,831  
FILING DATE: 04-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Churchill, Margaret A.  
REGISTRATION NUMBER: 39,944  
REFERENCE/DOCKET NUMBER: 1279-194C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 213/680-9200  
TELEFAX: 213/680-4518  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 503 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-582-776C-8

Query Match 100.0%; Score 24; DB 3; Length 503;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
DB 376 VSGSV 380

RESULT 9  
US-08-434-831B-8  
Sequence 8, Application US/08434831B  
Patent No. 6133905  
GENERAL INFORMATION:  
APPLICANT: Lipkin, W. I.  
APPLICANT: Briese, Thomas  
APPLICANT: Kliche, Stefanie  
APPLICANT: Schneider, Patrick A.  
APPLICANT: Stitz, Lothar  
APPLICANT: Schneemann, Anette  
TITLE OF INVENTION: Borina Disease Viral Sequences,  
TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
TITLE OF INVENTION: System Diseases  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski, L.L.P.  
STREET: 865 South Figueroa Street, 29th Floor  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90017-2571  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0  
SOFTWARE: ASCII (DOS) TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/434,831B  
FILING DATE: 04-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/369,822  
FILING DATE: 06-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Churchill, Margaret A. (Ph.D.)  
REGISTRATION NUMBER: 39,944  
REFERENCE/DOCKET NUMBER: 1279-194C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 213/680-9200  
TELEFAX: 213/680-4518  
INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:  
LENGTH: 503 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-434-831B-8

Query Match 100.0%; Score 24; DB 3; Length 503;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
DB 376 VSGSV 380

RESULT 10  
US-07-932-915-2  
Sequence 2, Application US/07932915  
Patent No. 5672486  
GENERAL INFORMATION:  
APPLICANT: Soullilou, Jean-Paul  
TITLE OF INVENTION: Protein Poly ligands joined To A Stable Protein  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum  
STREET: 5 Palo Alto Square, Suite 400  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/932,915  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/646,875  
FILING DATE: 28-JAN-1991  
APPLICATION NUMBER: US 07/575,394  
FILING DATE: 23-AUG-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland Ph.D., Bertram I.  
REGISTRATION NUMBER: 20,015  
REFERENCE/DOCKET NUMBER: ATLA-001/01US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-494-7622  
TELEFAX: 415-857-0663  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 504 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-932-915-2

Query Match 100.0%; Score 24; DB 1; Length 504;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
DB 206 VSGSV 210

RESULT 11  
PCT-US91-05826-2  
Sequence 2, Application PC/TUS9105826  
GENERAL INFORMATION:  
APPLICANT: Soullilou, Jean-Paul  
TITLE OF INVENTION: Protein Poly ligands joined To A Stable Protein  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bertram I. Rowland, Ph.D.  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/05826  
FILING DATE: 19910822  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/575,394  
FILING DATE: 29-AUG-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland Ph.D., Bertram I.  
REGISTRATION NUMBER: 20,015  
REFERENCE/DOCKET NUMBER: FP5352-1/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-494-8771  
TELEFAX: 415-781-1989  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 504 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-05826-2

Query Match 100.0%; Score 24; DB 5; Length 504;  
Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGSV 5  
DB 206 VSGSV 210

RESULT 12  
US-08-745-934-3  
Sequence 3, Application US/08745934  
Patent No. 5861496  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Hawkins, Phillip R.  
TITLE OF INVENTION: HUMAN SQUALENE EPOXIDASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/745,934
FILING DATE: Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0151 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 855-0555
TELEFAX: (415) 845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 572 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1217593
US-08-745-934-3

Query Match          100.0%; Score 24; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGSV 5
    |||||
Db 127 VGSV 131

RESULT 13
US-09-147-009-6
Sequence 6, Application US/09147009
Patent No. 6153815
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Process for Raising Squelene Levels in Plants
TITLE OF INVENTION: and DNA Sequences Used Therefor
NUMBER OF SEQUENCES: 11
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/147,009
FILING DATE:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 572 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Mus Musculus
STRAIN: B6CBA
DEVELOPMENTAL STAGE: 6-8 weeks
TISSUE TYPE: liver
IMMEDIATE SOURCE:
LIBRARY: Lambda ZAP vector Stratagene catalog #935302
CLONE: PMNSE-17
PUBLICATION INFORMATION:
AUTHORS: Kosuga, K.
AUTHORS: Hata, S.
AUTHORS: Osumi, T.
AUTHORS: Sakakibara, J.
AUTHORS: Ono, T.

```

```

TITLE: Nucleotide sequence of a cDNA for mouse
TITLE: squelene epoxidase
JOURNAL: Biochim. Biophys. Acta
VOLUME: 1260
ISSUE: 3
PAGES: 345-348
DATE: 1995
US-09-147-009-6

Query Match          100.0%; Score 24; DB 4; Length 572;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGSV 5
    |||||
Db 127 VGSV 131

RESULT 14
US-08-050-684-2
Sequence 2, Application US/08050684
Patent No. 5550221
GENERAL INFORMATION:
APPLICANT: Johann Dr., Stephen V.
APPLICANT: Van Zeijl Dr., Marja
APPLICANT: O'Hara Dr., Bryan M.
TITLE OF INVENTION: Amphotropic Virus Receptor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSER: American Cyanamid Company
STREET: 1937 West Main Street
CITY: Stamford
STATE: CT
COUNTRY: United States of America
ZIP: 06904-0060
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/050,684
FILING DATE: 16-APR-1993
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Lowney Dr., Karen A
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 31937-00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 203-321-2361
TELEFAX: 203-321-2971
TELEX: 710-474-4059
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-050-684-2

Query Match          100.0%; Score 24; DB 1; Length 652;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGSV 5
    |||||
Db 37 VGSV 41

RESULT 15
US-08-582-719-2

```

```
; Sequence 2, Application US/08582719
; Patent No. 5633348
; GENERAL INFORMATION:
; APPLICANT: Johann Dr., Stephen V.
; APPLICANT: Van Zeijl Dr., Marja
; APPLICANT: O'Hara Dr., Bryan M.
; TITLE OF INVENTION: Ampholropic Virus Receptor
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: 1937 West Main Street
; CITY: Stamford
; STATE: CT
; COUNTRY: United States of America
; ZIP: 06904-0060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/582,719
; FILING DATE: 04-JAN-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/050,684
; FILING DATE: 16-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney Dr., Karen A
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 31937-00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203-321-2361
; TELEFAX: 203-321-2971
; TELEK: 710-474-4059
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-582-719-2

Query Match          100.0%; Score 24; DB 1; Length 652;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VGSgv 5
        |||||
DB      37 VGSgv 41
```

Search completed: September 6, 2001, 16:39:38  
Job time: 127 sec





PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
PT having 2 catalytic aspartic residues and substrate binding cleft, used  
PT to treat Alzheimer's disease by blocking amyloid precursor protein  
PT cleavage -  
XX  
PS Example 4; Page 33; 86pp; English.  
XX  
CC The present sequence is given in a specification relating to an inhibitor  
CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
CC active site, which is defined by the presence of two catalytic aspartic  
CC residues and a substrate binding cleft. The inhibitor is useful for  
CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
CC screens for individuals with a genetic predisposition to Alzheimer's  
CC disease. The inhibitor is useful as a reagent for specifically binding to  
CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
CC isolation, purification and characterisation.  
SQ Sequence 5 AA;  
  
Query Match 100.0%; Score 24; DB 22; Length 5;  
Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VSGSV 5  
      |||||  
DB 1 vsgsv 5  
  
RESULT 2  
AAB61344  
ID AAB61344 standard; peptide; 5 AA.  
XX  
AC AAB61344;  
XX  
DT 02-APR-2001 (first entry)  
XX  
DE Memapsin 2 substrate #3.  
XX  
KW Memapsin 2; catalyst; Alzheimer's.  
XX  
OS Synthetic.  
XX  
PM WO200100663-A2.  
PD  
XX 04-JAN-2001.  
PF  
XX 27-JUN-2000; 2000WO-US17661.  
PR  
XX 28-JUN-1999; 99US-0141363.  
PR 30-NOV-1999; 99US-0168060.  
PR 25-JAN-2000; 2000US-0177836.  
PR 27-JAN-2000; 2000US-0178368.  
PR 08-JUN-2000; 2000US-0210292.  
XX  
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
XX  
PI Tang JUN, Lin X, Koelsch G;  
XX  
DR WPI; 2001-102865/11.  
XX  
PT Purified recombinant catalytically active memapsin 2, used to screen  
PT inhibitors of it, which are used to treat and prevent Alzheimer's  
PT disease -  
XX  
PS Example 4; Page 33; 86pp; English.  
XX  
CC The present invention relates to a purified recombinant  
CC catalytically active memapsin 2. The invention may be used for  
CC isolating inhibitors which are used to treat or prevent  
CC Alzheimer's disease. The invention may also be used to screen  
CC for individuals more genetically prone to develop Alzheimer's  
CC disease.

XX  
SQ Sequence 5 AA;  
  
Query Match 100.0%; Score 24; DB 22; Length 5;  
Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VSGSV 5  
      |||||  
DB 1 vsgsv 5  
  
RESULT 3  
AAB66583  
ID AAB66583 standard; Peptide; 7 AA.  
XX  
AC AAB66583;  
XX  
DT 12-APR-2001 (first entry)  
XX  
DE Synthetic memapsin 2 substrate peptide #2.  
XX  
KW memapsin 2; nootropic; neuroprotective; amyloid precursor protein;  
KW APP; memapsin 2 inhibitor; Alzheimer's disease.  
XX  
OS Synthetic.  
XX  
PM WO200100665-A2.  
PD  
XX 04-JAN-2001.  
PF  
XX 27-JUN-2000; 2000WO-US17742.  
PR  
XX 28-JUN-1999; 99US-0141363.  
PR 30-NOV-1999; 99US-0168060.  
PR 25-JAN-2000; 2000US-0177836.  
PR 27-JAN-2000; 2000US-0178368.  
PR 08-JUN-2000; 2000US-0210292.  
XX  
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
XX  
PI (UNIL) UNIV ILLINOIS FOUND.  
XX  
PI Tang JUN, Hong L, Ghosh AK;  
XX  
DR WPI; 2001-137933/14.  
XX  
XX  
XX Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
XX having 2 catalytic aspartic residues and substrate binding cleft, used  
XX to treat Alzheimer's disease by blocking amyloid precursor protein  
XX cleavage -  
XX  
PS Example 4; Page 33; 86pp; English.  
XX  
CC The present sequence is given in a specification relating to an inhibitor  
CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
CC active site, which is defined by the presence of two catalytic aspartic  
CC residues and a substrate binding cleft. The inhibitor is useful for  
CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
CC screens for individuals with a genetic predisposition to Alzheimer's  
CC disease. The inhibitor is useful as a reagent for specifically binding to  
CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
CC isolation, purification and characterisation.  
SQ Sequence 7 AA;  
  
Query Match 100.0%; Score 24; DB 22; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 VSGSV 5  
      |||||



Db 1 vgsqv 5

RESULT 4  
ID AAB61345 standard; peptide; 7 AA.  
XX  
AC AAB61345;  
XX  
DT 02-APR-2001 (first entry)  
XX  
DE Memapsin 2 substrate #4.  
XX  
KW Memapsin 2; catalyst; Alzheimer's.  
XX  
OS Synthetic.  
XX  
PN WO200100663-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 27-JUN-2000; 2000WO-US17661.  
XX  
PR 28-JUN-1999; 9905-0141363.  
PR 30-NOV-1999; 9905-0168060.  
PR 25-JAN-2000; 2000US-0177836.  
PR 27-JAN-2000; 2000US-0178368.  
PR 08-JUN-2000; 2000US-0210292.  
XX  
PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
XX  
PI Tang JTN, Lin X, Koelsch G;  
XX  
DR WPI: 2001-102885/11.  
XX  
PT Purified recombinant catalytically active memapsin 2, used to screen  
XX inhibitors of it, which are used to treat and prevent Alzheimer's  
XX disease -  
XX  
PS Example 4; Page 33; 86pp; English.  
XX  
CC The present invention relates to a purified recombinant  
XX catalytically active memapsin 2. The invention may be used for  
XX isolating inhibitors which are used to treat or prevent  
XX Alzheimer's disease. The invention may also be used to screen  
XX for individuals more genetically prone to develop Alzheimer's  
XX disease.  
XX  
SQ Sequence 7 AA:

Query Match 100.0%; Score 24; DB 22; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGSQV 5  
Db 1 vgsqv 5

RESULT 5  
ID AAY23474 standard; Peptide; 8 AA.  
XX  
AC AAY23474;  
XX  
DT 02-SEP-1999 (first entry)  
XX  
DE V beta 6 clone found in MS patients after vaccination with TCR.  
XX  
KW Vaccine; T cell receptor; TCR; T cell; V beta 6.2/3; V beta 6/5;  
XX V beta 6.7; V beta 2; V beta 5/1; V beta 7; V beta 13; V beta 6;  
XX multiple sclerosis.

XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO9927957-A1.  
XX  
PD 10-JUN-1999.  
XX  
PF 03-DEC-1997; 97WO-US23147.  
XX  
PR 03-DEC-1997; 97WO-US23147.  
XX  
PA (IMMU-) IMMUNE RESPONSE CORP.  
XX (KIMM-) KIMMEL CANCER CENT SIDNEY.  
XX  
PI Brostoff SW, Carlo DJ, Gold DP, Smith LR, Wilson DB;  
XX  
DR WPI: 1999-404801/34.  
XX  
PT T0 cell receptor peptide-derived vaccines  
XX  
PS Example 11; Page 88; 104pp; English.  
XX  
CC The specification describes vaccines which comprise immunologically  
XX effective amounts of T cell receptor (TCR) peptides. The TCRs are  
XX present on the surface of T cells. The TCRs are chosen from V beta  
XX 6.2/3, V beta 6/5, V beta 6.7, V beta 2, V beta 5/1, V beta 7 or V beta  
XX 13. The V beta TCR peptide-based vaccines are useful for prevention or  
XX treatment of multiple sclerosis (MS). The presence of V beta 6.7 appears  
XX to be particularly associated with multiple sclerosis and can be used  
XX to determine an individual's susceptibility to multiple sclerosis.  
XX Vaccinating, rather than passively administering heterologous  
XX antibodies, allows the host's own immune system to mobilize and suppress  
XX auto aggressive T cells. Therefore, the suppression is persistent and  
XX may involve any and all immunological mechanisms in effecting that  
XX suppression. Such a multi-faceted response is more effective than  
XX the uni-dimensional suppression achieved by passive administration of  
XX monoclonal antibodies or extant-derived regulatory T cell clones.  
XX AAY23387-Y23480 represent peptides derived from TCR V beta 6 clones  
XX found in the cerebrospinal fluid (CSF) of MS patients, after vaccination  
XX with V beta 6.  
XX  
SQ Sequence 8 AA:

Query Match 100.0%; Score 24; DB 20; Length 8;  
Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGSQV 5  
Db 3 vgsqv 7

RESULT 6  
ID AAB65588 standard; Peptide; 10 AA.  
XX  
AC AAB65588;  
XX  
DT 12-APR-2001 (first entry)  
XX  
DE Synthetic memapsin 2 substrate peptide.  
XX  
KW memapsin 2; neurotrophic; neuroprotective; amyloid precursor protein;  
XX APP; memapsin 2 inhibitor; Alzheimer's disease.  
XX  
OS Synthetic.  
XX  
PN WO200100665-A2.  
XX  
PD 04-JAN-2001.  
XX

PF 27-JUN-2000; 2000WO-US17742.  
 XX  
 XX 28-JUN-1999; 99US-0141363.  
 PR 30-NOV-1999; 99US-0168060.  
 PR 25-JAN-2000; 2000US-0177836.  
 PR 27-JAN-2000; 2000US-0178368.  
 PR 08-JUN-2000; 2000US-0210292.  
 XX  
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA (UNIT ) UNIV ILLINOIS FOUND.  
 XX  
 PI Tang JYN, Hong L, Ghosh AK;  
 XX  
 DR WPI; 2001-137933/14.  
 XX  
 PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2  
 PT having 2 catalytic aspartic residues and substrate binding cleft, used  
 PT to treat Alzheimer's disease by blocking amyloid precursor protein  
 PT cleavage -  
 PS  
 PS Example 4; Page 33; 86pp; English.  
 XX  
 CC The present sequence is given in a specification relating to an inhibitor  
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2  
 CC active site, which is defined by the presence of two catalytic aspartic  
 CC residues and a substrate binding cleft. The inhibitor is useful for  
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in  
 CC screens for individuals with a genetic predisposition to Alzheimer's  
 CC disease. The inhibitor is useful as a reagent for specifically binding to  
 CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2  
 CC isolation, purification and characterisation.  
 CC  
 XX  
 SQ Sequence 10 AA:  
 XX  
 Query Match 100.0%; Score 24; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSV 5  
 |||||  
 DB 1 vsgsv 5

XX  
 DR WPI; 2001-102885/11.  
 XX  
 XX Purified recombinant catalytically active memapsin 2, used to screen  
 PT inhibitors of it, which are used to treat and prevent Alzheimer's  
 PT disease -  
 PT  
 XX  
 XX Disclosure; Page 86; 86pp; English.  
 XX  
 PS  
 CC The present invention relates to a purified recombinant  
 CC catalytically active memapsin 2. The invention may be used for  
 CC isolating inhibitors which are used to treat or prevent  
 CC Alzheimer's disease. The invention may also be used to screen  
 CC for individuals more genetically prone to develop Alzheimer's  
 CC disease.  
 CC  
 XX  
 SQ Sequence 10 AA:  
 XX  
 Query Match 100.0%; Score 24; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 VSGSV 5  
 |||||  
 DB 1 vsgsv 5

RESULT 8  
 ID AAY41938 standard; Peptide; 16 AA.  
 XX  
 AC AAY41938;  
 XX  
 DT 09-DEC-1999 (first entry)  
 XX  
 DE Rheumatoid arthritis diagnostic protein isoform peptide #89.  
 XX  
 DE Human; Rheumatoid arthritis; RA; diagnosis; RPI; RADP; detection;  
 KW Rheumatoid arthritis diagnostic feature; EPR; synovial fluid;  
 KW Rheumatoid arthritis diagnostic protein isoform; screening;  
 KW expression reference protein isoform; prognosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9947925-A2.  
 XX  
 PD 23-SEP-1999.  
 XX  
 PF 15-MAR-1999; 99WO-GB00763.  
 XX  
 PR 13-MAR-1998; 98GB-0005477.  
 XX  
 PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 XX  
 PI Parekh RB, Patel TP, Townsend RR;  
 XX  
 DR WPI; 1999-571871/48.  
 XX  
 PT Diagnosis of human rheumatoid arthritis by two-dimensional  
 PT electrophoresis -  
 PT  
 XX  
 PS Disclosure; Page 19; 157pp; English.  
 XX  
 CC A method has been developed for the diagnosis of human rheumatoid  
 CC arthritis (RA) using two-dimensional electrophoresis to generate a  
 CC two-dimensional array of features. The method can be used for screening,  
 CC diagnosis and prognosis of RA in a subject or for monitoring the effect  
 CC of an anti-RA drug or therapy administered to a subject. The method  
 CC comprises: (a) analysing a sample of serum or plasma and optionally  
 CC synovial fluid by two-dimensional electrophoresis, to generate a two-  
 CC dimensional array of features; (b) identifying at least one chosen  
 CC feature whose relative abundance correlates with the presence or absence

CC of RA, and (c) comparing the abundance of each chosen feature in the  
CC sample with the abundance of that chosen feature in serum or plasma from  
CC one or more persons without RA, where the relative abundance of the  
CC chosen feature or features in the sample indicates the presence or  
CC absence of RA in the subject. The method can also be used in clinical  
CC studies for testing drugs for therapy of RA, for purification of RA-  
CC diagnostic protein isoforms (RPIs), and for production of antibodies to  
CC RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify  
CC compounds that promote or inhibit their activity, which are then used as  
CC RA drugs. Nucleic acid encoding RADPs can be used in gene therapy  
CC protocols. AAY41844 to AAY42100 represent RPI peptides, AAY42101 to  
CC AAY42103 represent expression reference protein isoform peptides and  
CC AA255066 to AA255068 represent degenerate probes for RPIs, which are  
CC used in the exemplification of the present invention.

**SQ Sequence 16 AA;**

Query Match	100.0%	Score	24	DB	20	Length	16
Best Local Similarity	100.0%	Pred. No.	80				
Matches	5	Conservative	0	Mismatches	0	Indels	0
						Gaps	0

QY	1	VGSGV	5
Db	2	VGSGV	6

## RESULT 9

AAV14046.  
ID AAV14046 standard; peptide; 20 AA

AC AAY14046;

DT 15-JUL-1999 (first entry)

DE N-terminus of D-sorbitol dehydrogenase.

KW D-sorbitol dehydrogenase; L-sorbose; 2-keto-L-gulonate precursor

L-ascorbic acid production; N-terminus.

OS *Glucanobacter oxydans*.

PN W09920763-A1.

PD 29-APR-1999.

PF 13-OCT-1998; 98WO-JP04612.

PR 17-OCT-1997; 97JP-0285280.

PA (FUJI ) FUJISAWA PHARM CO LTD.

PI Ishii Y, Noguchi Y, Salto Y, Soeda S, Yoshikawa K;

DR WPI; 1999-302741/25.

PT Gene group for D-sorbitol dehydrogenase, useful for simple large-scale production of L-sorbose or 2-keto-L-gulononic acid as precursor for L-ascorbic acid

PS Claim 1; Page 40; 83pp; Japanese

CC This sequence represents a fragment of the D-sorbitol dehydrogenase  
CC of the invention. Cells transformed with a vector containing DNA encoding  
CC the dehydrogenase can be used to produce L-sorbitose or 2-keto-L-gulonate  
CC acid as precursor for small large-scale L-ascorbic acid production.

**SQ Sequence . 20 AA;**

Query Match	100.0%;	Score 24;	DB 20;	Length 20;
Best Local Similarity	100.0%;	Pred. No. 1e+02;		
Matches	5;	Conservative 0;	Mismatches 0;	Indels 0;
			Gaps	0

Qy	1	VGSGV	5
Db	13	vgsgv	17

RESULT 10

ID AAB27909 standard; Protein; 72 AA.

AC AAB27909;

DT 02-FEB-2001 (first entry)

DE Human secreted protein' SEQ ID NO: 63.

KM Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KM vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder  
KM neurological disease; infection; human; secreted protein.

OS Homo sapiens.

PN WO2000055171-A1

PD 21-SEP-2000.

PF 09-MAR-2000; 2000WO-US06043

PR 12-MAR-1999; 99US-0124146

XX

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM, Komatsoulis G;

DR WPI; 2000-638174/61.

DR N-PSDB; AAC59051.

PR Isolated nucleic acid molecule encoding a human secreted protein is  
 PR used in preventing, treating or ameliorating a medical condition -  
 XX  
 PS Claim 11, Page 365, 438pp. English.

CC Sequences AAB27907-B27956 represent the amino acid sequences of 50  
CC human secreted proteins encoded by the genes AAC59049-C59098. The genes  
CC and proteins are useful for preventing, ameliorating or treating medical  
CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
CC a range of human tissues disclosed in the specification. The nucleic  
CC acids, proteins, antibodies and (ant)agonists are useful in the  
CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
CC Crohn's disease, multiple sclerosis, Rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.

50 Sequence 72 AA

Query Match	100.0%	Score 24	DB 21	Length 72
Best Local	Similarity 100.0%	Pred. No. 3.5e+02		
Matches	5	Conservative 0	Mismatches 0	Indels 0
				Gaps 0

Qy	1	VCSCV	5
Db	65	vgsgv	65

```

RESULT 11
AAR22386
ID AAR22386 standard; Protein; 87 AA.
XX
AC AAR22386;
XX
DT 17-AUG-1992 (first entry)
XX
DE Antigen tc-32c.
XX
Oocysts; antiserum; sporozoite; E. maxima.
XX
KW Elmeria tenella.
XX
OS WO9204461-A.
XX
PN 19-MAR-1992.
XX
PD 05-SEP-1991; 91WO-US06431.
XX
PF 12-SEP-1990; 90US-0581694.
XX
PR (GENEX-) GENEX CORP.
XX
PA (FARN) HORCHST AG.
XX
PI Jacobson JW, Strausberg RL, Wilson SD, Pope SH, Strausberg SL;
PI Raether W;
XX
DR MPI: 1992-114366/14.
XX
N-PSDB; AAQ23086.
XX
PT Vaccine against avian coccidiosis - comprising recombinant
PT microorganisms expressing them
XX
PS Claim 20; Page 66 + Fig 20; 94pp; English.
XX
XX To identify antigens of E. tenella, expression libraries were
CC prep. in the lambda vector, lambda gill, using cDNA prep. from
CC polyA mRNA isolated from E. tenella oocysts. The cDNA expression
CC library was screened with rat anti-E. maxima sporozoite immune
CC serum. The cDNA inserts from the positive clones were cloned into
CC bacteriophage M13 and subjected to sequence analysis.
CC E. tenella antigen tc-32c (8.8 kD) was identified. Antigen tc-32c
CC consists of 261 bp of ORF encoding an antigen fragment.
XX
SQ Sequence 87 AA;
XX
Query Match 100.0%; Score 24; DB 13; Length 87;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VGSgv 5
DB 40 VgsGV 44
XX
RESULT 12
AAW42024
ID AAW42024 standard; Protein; 95 AA.
XX
AC AAW42024;
XX
DT 28-MAY-1998 (first entry)
XX
DE Clone K330_2 protein.
XX
KW K330_2 clone; cDNA library; murine bone marrow; cell proliferation;
KW cell differentiation.
XX
OS Mus sp.

```

```

XX
PN US5708157-A.
XX
PD 13-JAN-1998.
XX
PF 26-JUL-1996; 96US-0686878.
XX
PR 26-JUL-1996; 96US-0686878.
XX
PA (GEMV ) GENETICS INST INC.
XX
PI (JACO/) JACOBS K.
XX
PI Evans C, Jacobs K, LaVallie ER, McCoy JM, Merberg D;
PI Racile LA, Spaulding V, Treacy M;
XX
DR MPI: 1998-109352/10.
XX
N-PSDB; AAV09111.
XX
PT Human testes derived cDNA and protein(s) - may be useful for
PT laboratory use, or for stimulating cell proliferation and
PT differentiation
XX
PS Disclosure; Column 67-68; 48pp; English.
XX
XX This sequence is the protein for the clone K330_2, which is 95 amino
XX acids in length, and is isolated from the murine bone marrow cDNA
XX library. When the K330_2 sequence was searched in the Genbank database,
XX the clone demonstrated at least some homology with an EST identified as
XX "y182c07.s1 Homo sapiens cDNA clone 28939 3'".
XX
SQ Sequence 95 AA;
XX
Query Match 100.0%; Score 24; DB 19; Length 95;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VGSgv 5
DB 73 VgsGV 77
XX
RESULT 13
AAV08632
ID AAV08632 standard; Protein; 95 AA.
XX
AC AAV08632;
XX
DT 10-AUG-1999 (first entry)
XX
DE Murine secreted protein K330_3.
XX
KW Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KW bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KW cell proliferation; cell differentiation; suppressor; tumour inhibition;
KW haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;
KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW cadherin; tumour invasion suppressor; gene therapy; tissue growth.
XX
OS Mus musculus.
XX
PN WO9926972-A1.
XX
PD 03-JUN-1999.
XX
PF 17-NOV-1998; 98WO-US24614.
XX
PR 20-OCT-1998; 98US-0175928.
XX
PR 21-NOV-1997; 97US-0976110.
XX
PR 18-MAY-1998; 98US-0080478.
XX
PA (GEMV ) GENETICS INST INC.
XX

```